Database Systems Laboratory

Subhash Bhalla
Professor

Yutaka Watanobe
Assistant Professor

Wanming Chu
Assistant Lecturer

Junya Terazono
Assistant Lecturer

Information systems in transportation, health-care and public utility services depend on large scale data management systems. Research activity in Database Systems is focused on broadening their range of applicability. It is also focused on improving the performance of domain specific applications.

Large applications for Services use database systems as a basic part for web data resources. This year the laboratory organized an International workshop with many invited distinguished researchers during 25-27 March 2013. The proceedings have been published for the 8th workshop on "Databases in Networked Information System (DNIS)" by Springer-Verllag in the Lecture Notes in Computer Science series (LNCS) in Volume 7813. The delivered lecture and manuscripts are being utilized to develop the state-of-the-art lectures on current research problems. These created a focused view on new research problems. many current aspects of web related research activity were discussed at the workshop. In addition to DNIS 2013 workshop, the laboratory organized an International conference with many invited distinguished researchers during 24 - 26 December 2012. The proceedings have been published for the conference (Ist BDA 2012) by Springer-Verllag in the Lecture Notes in Computer Science series (LNCS) in Volume 7678. The delivered lecture created a focused view on new research problems.

Most of the advances in techniques concentrate on capturing more meaning within data. A number of researchers are actively developing improved data management strategies using Business Intelligence and data. This provides a challenging area for study. The domains of activity include :

- developing new user interfaces and query languages for skilled and semi-skilled users in health-care;

- developing infrastructures for computing facilities for cloud computing;
- supporting mobile computing applications, and
- designing new data models and asynchronous computation models for transaction and services.

In addition to complexity in features such as multiple attributed data, many new types of raw data are emerging that need to be captured by DBMSs for information extraction. Many research efforts are being made to make suitable Object-Relational architectures for spatial databases.

The Database Systems Laboratory has research activity concentrating on data modeling as well as transaction processing activity.

Data Modeling for Spatial Objects

Information Processing Systems of future will be a combination of integrated components. There will be components for intelligent problem solving, or decision making, components for specialized data processing and components for shared information management. The applications will utilize a shared base of information. Some examples are - business automation, industrial automation, computer-aided design and manufacture, and cartography.

Work-flow and Web Services

Traditional approaches to transaction management introduce elements of unpredictable delays during transaction processing. Thus, making these not suitable for adoption in new application environments. The techniques for time-critical transactions are applicable to Mobile databases and multimedia databases.

Healthcare Studies

The goal of this research is to study the Standardized Electronic Health Records (EHRs) databases. It is a temporal computational system with the ability to process large volume of information. Such system will prove useful in various areas of information technology such as online healthcare agencies. The modeling considers the complementary points of view:

1. EHRs data mining approach to address the epidemic studies,

2. An approach that involves the user in the modeling process.

3. Query Language with reference to user skills

Prototype systems to access dynamic contents through web based information systems are in progress. These emulate mobile e-commerce activity in banking and Geographic Information Systems, for test and studies. The test protypes have been evolved based on research on new easy-to-use search and also new query language interfaces.

Summary of Achievement

Refereed Journal Papers

Unrefereed Papers

[terazono-01:2012] H. Yoshida, M. Kondo, K. Ando, K. Usuki, Y. Sato, H. Demura, J. Terazono, N. Toshima, Y. Nakamura, S. Nozaki, A. Mizutani, R. Mochiji, A. Morimoto, Y. Yamamoto, and J. Watanabe. Activities of the Liaison Group of Fukushima for 2012 Annular Eclipse. *Journal of Center for Regional Affairs, Fukushima University*, 24(1):114–135, 2012.

This is an activity record of the Liaison Group of Fukushima for 2012 Annular Eclipse, one of the most important space-related event in Japan on 2012. This group is formed by university researchers, museum specialists, amateur and professional astronomer and exchanged information on the Annular Eclipse including risk of viewing and related projects. The group worked to utilize this special opportunity for scientific outreach and education for young people especially for elementary school pupil and junior/senior high school students.

Refereed Proceeding Papers

[bhalla-04:2012] A. Madaan, W. Chu, D. Yaginoma, and S. Bhalla. Quasi-Relational Query Language Interface for Persistent Standardized EHRs: Using NoSQL Databases. In A. Madaan, S. Kikuchi, and S. Bhalla, editors, *8th International Workshop on Databases in Networked Information Systems*, pages 84–97, Germany, March 2013. Lecture Notes in Computer Science Series, Springer Verlag.

Interoperability of health data for information exhange is an area of growing concern. Various new standards such as CEN 13606, HL7 and OpenEHR have been proposed. The OpenEHR standard provides a Standardized Electronic Health Records (EHRs) schema using dual-level modelling for information exchange. The complex structured EHRs and the archetypes form the domain knowledge of the model. It gives rise to the issue of efficient and scalable persistence mechanism for these standardized EHRs. Further, it is desirable to support in-depth query-ability on them. The standardized EHRs database can support a wide range of user queries. In this paper, a persistence mechanism using a NoSQL database for storing the standardized EHRs has been

proposed. Further, a high-level QBE-like AQBE (Archetype based Query-By-Example) has been evolved for the EHRs data repository.

[bhalla-05:2012] A. Yasir, M. K. Swamy, P. K. Reddy, and S. Bhalla. Enhanced Query-By-Object Approach for Information Requirement Elicitation in Large Databases. In V. Bhatnagar and S. Srinivasa, editors, *Ist International Conference on Big Data Analytics*, pages 26–41, Germany, December 2012. Lecture Notes in Computer Science Series, Springer Verlag.

Information Requirement Elicitation (IRE) recommends a framework for developing interactive interfaces, which allow users to access database systems without having prior knowledge of a query language. An approach called ‘ Query-by-Object ’ (QBO) has been proposed in the literature for IRE by exploiting simple calculator like operations. However, the QBO approach was proposed by assuming that the underlying database is simple and contains few tables of small size. In this paper, we propose an enhanced QBO approach called Query-by-Topics (QBT), for designing calculator like user interfaces for large databases. We use methodologies for clustering database entities and discovering topical structures to represent objects at a higher level of abstraction. The QBO approach is then enhanced to allow users to query by topics (QBT). We developed a prototype system based on QBT and conducted experimental studies to show effectiveness of the proposed approach.

[yutaka-04:2012] Yuya Watanabe Yutaka Watanobe, Nikolay N. Mirenkov. AIDA compiler: a code synthesizer from programs in pictures. In *The Joint International Conference on Human-Centered Computer Environments HCCE2012*, pages 76–83, 2012.

AIDA is a language for programming (modeling and documenting) in pictures within F-modeling environment where pictures and moving pictures are used as super-charactersfor representing features of computational algorithms, corresponding application models and possible related documentation. Generic pictures of the algorithmic super-characters are used to compose compound pictures defining algorithmic steps. The generic and compound pictures, as well as their series are acquired in special galleries of an open type where supportive pictures of various annotations are also included. These acquisitions allow very compact algorithmic specifications from which a set of program slices with different levels of details are synthesized. The lowest level slice is a C++ program. The F-modeling environment provides a set of editors for different views and annotations. In this paper, a structure of the synthesizer

is presented and a distribution of its functions among the environment subsystems is provided. A special attention is payed to internal objects, template programs and algorithms supporting the C++ program generation.

[yutaka-05:2012] Yutaka Watanobe Ryuichi Oka Takeshi Sasaki, Yuichi Yaguchi. Extracting a Spatial Ontology from a Large Flickr Tag Dataset. In *The 4th International Conference on Awareness Science and Technology*, pages 91–97, 2012.

We propose an easy framework for automatically constructing spatial ontologies that locate related concepts together in a space. The conventional graph representation is strong in showing direct relationships between entities, but it is difficult to process its topology when extracting features from the network, because similarity between networks is not well determined. Spatial ontologies are easy to cluster and classify according to the similarities or relationships between entities. We propose a method for creating a spatial ontology called "Associated Keyword Space" and apply it to 0.4M tag words collected from more than 1M images in Flickr. Tags in Flickr have many unknown word tags, but the spatial ontology can explain the clusters of meaning including unknown word tags. The results show that these unknown word tags can be found from neighbor tags that have clear meanings. As a result, an "area ontology" can be explained from the spatial ontology.

Unrefereed Papers

[terazono-02:2012] J. Terazono, R. Nakamura, S. Kodama, N. Yamamoto, N. Hirata, H. Demura, Y. Ogawa, and T. Sugawara. WISE-CAPS: Data Archiving, Browsing and Analyzing Environment for Lunar and Plaentary Data. In *Planetary Data Workshop, USGS Flagstaff*, page none, June 2012.

We are now constructing a Web-GIS based framework called "WISE-CAPS," a virtual research environment to share and browse the data under secured data access. This presentation will address on current implementation status and future prospective.

Books

[bhalla-06:2012] A. Madaan, S. Kikuchi, and S. Bhalla, editors. *Databases in Networked Information Systems*, volume 7813 of *Lecture Notes in Computer Science*. Springer-Verlag, Germany, March 2013.

Grants

[yutaka-06:2012] Yutaka Watanobe. 次世代プログラミング言語ＡＩＤＡ, 2012.

プログラマが思い描くモデル・アルゴリズムをより直接的に記述し、生産性と信頼性が高いソフトウェア開発プロセスを支援する次世代のプログラミング言語ＡＩＤＡの研究開発を行う。ＡＩＤＡはオープンな言語要素を拡充することによって知識の蓄積を行うユニバーサルな言語であるため、本課題において企業等が所有する実用的なアプリケーション（解析コード等）をＡＩＤＡにより実装する。さらに、言語要素や開発環境に関してユーザビリティテストを継続して行い、関連企業や機構への技術移転と実用化を目指す。

Academic Activities

[yutaka-07:2012] Yutaka Watanobe, 2012.

Reviewer

[yutaka-08:2012] Yutaka Watanobe, 2012.

Reviewer

[yutaka-09:2012] Yutaka Watanobe, 2013.

Reviewer

[yutaka-10:2012] Yutaka Watanobe, 2012.

Reviewer

Patents

[yutaka-11:2012] 渡部 有隆 ニコライ ミレンコフ, 吉岡 廉太郎. 編集支援プログラム、プログラム改変の支援方法、プログラム編集の支援方法およびコンピュータ処理方法, 2012 (approved).

Ph.D and Others Theses

[terazono-03:2012] Hiroki Akamatsu. Implementation of In-page Programming Environment in GIS, University of Aizu, 2013.

Thesis Advisor: J. Terazono

[terazono-04:2012] Susumu Ito. Creation of Web API for Lunar and Plaentary GIS, University of Aizu, 2013.

Thesis Advisor: J. Terazono

[terazono-05:2012] Yoshihide Okazaki. Tsukiyomi II: Enhancements of Lunar Photograph Search System, University of Aizu, 2013.

Thesis Advisor: J. Terazono

[terazono-06:2012] Noritaka Shimizu. Application of Contents Control Scheme for Large-Scaled Website, University of Aizu, 2013.

Thesis Advisor: J. Terazono

[w-chu-02:2012] Masanari Hirokawa. Graduation Thesis: Search Interface for Large Wikipedia Dump, University of Aizu, 2012.

Thesis Advisor: Wanming Chu

[yutaka-12:2012] Yuya Watanabe. Filmi

cation of Methods: A Visual Programming Language Based on Super-characters, University of Aizu, 2012.

Filmi

cation of Methods (Fim) is a visual programming paradigm which data and knowledge are transformed into much more high-level information resources including models, codes, and doc- uments. The approach employs multiple views, pictures and moving pictures to represent features of algorithms and computational process. Generic pictures of the algorithmic super-characters are used to represent features of computational models and algorithms. Such super-characters are com- posed into a program. Many super-characters involved in the program are intuitively understandable and do not require serious effects for memorization. However, to enhance the comprehension of the super-characters and their compositions, a majority of super-characters for the program is supported by annotations for the explanation based on appropriate mediums such as texts, diagrams, images, movies, a scene of frames. The key features of visual programming is the use of visual techniques to specify a program, and a visual

programming language uses visual (or graphical) notations. The aim of the visual programming languages is to provide better communication between human and computer as well as between different programmers to improve human performance. So far, a number of visual programming languages and environments have been developed for different domains and in different approaches. They contribute to develop applications in speci

c domains. On the other hand, it is difficult to use them as general purpose languages because of their limitations, abstractness, size, the number of notations etc. In this paper, a novel approach for a visual programming language within the Fim technology is presented. The approach is to improve code understandability based on annotations for super- characters as well as to make it a general purpose language, which has much more expressive power based on existing libraries and external services while using much more simple language constructs. In addition, it supports special functions for facilitating coding through binding, spliced elements, folding/unfolding elements, etc. We provide basic concepts of the language, its grammar, environ- ment, library, algorithm of code generation, and editing operations. Case studies of different levels are also demonstrated.

[yutaka-13:2012] Kanto Nakayama. Data Visualization in *AIDA, University of Aizu, 2012.

This paper explains implementation of data visualization methods and the corresponding functions for *AIDA. *AIDA is a programming/modeling language representing algorithms by pictures and moving pictures. *AIDA and its modeling environment visualize methods as well as output results. Hence, libraries of the functions enable *AIDA users to visualize their computational result on various methods of visualization. Web interfaces implementing the libraries based on JavaScript and HTML5 are developed. Then, the libraries are implemented to the environment. The libraries package not only basic methods such as bars, pies, lines, but also particular methods such as chords, meshes for charts, and space structures such as graphs. The libraries support *AIDA user's simulation and experiment; moreover, they give the users cognitive enhancement and increase understandability about the corresponding models, simulations, and algorithms.

[yutaka-14:2012] Takeshi Sasaki. Extracting a Spatial Ontology from a Large Flickr Tag Dataset, University of Aizu, 2012.

We propose an easy framework for automatically constructing spatial ontologies that locate related concepts together in a space. The conventional graph representation is strong in showing direct relationships between entities, but it is

difficult to process its topology when extracting features from the network, because similarity between networks is not well determined. Spatial ontologies are easy to cluster and classify according to the similarities or relationships between entities. We propose a method for creating a spatial ontology called “Associated Keyword Space” and apply it to 0.4M tag words collected from more than 1M images in Flickr. Tags in Flickr have many unknown word tags, but the spatial ontology can explain the clusters of meaning including unknown word tags. The results show that these unknown word tags can be found from neighbor tags that have clear meanings. As a result, an “area ontology” can be explained from the spatial ontology.

[yutaka-15:2012] Tsukasa Arima. Applying Gamification to Online Programming Challenge Service, University of Aizu, 2012.

Gamification methods for online programming environment is presented. The gamification is the use of game mechanics for non-game services to encourage users to adopt the service. The online programming environment is a judging system (service) for programming assignments to support university students, high school students or other programmers. The service includes about 10,000 users and 1,000 problems and more than a half million submitted codes. In this paper, gamification for the programming environment is applied based on a framework for the gamification including task, reward and social interchange. As concrete elements for the programming environment, visualization, purpose element, social action and play cycle are considered.

[yutaka-16:2012] Kandai Saito. Filmification of Methods: A Modeling Environment Based on Super-charcters, University of Aizu, 2012.

A modeling environment based on algorithminc supercharcters is presented. Filmification of Methods is a programming paradigm where data/knowlage is transformed into more high-level informarion resources. The argorithmic supercharacters are defined by intelligible pictures which represent computational models and algorithms. They are supported by a set of annotations which exlpain meaning of the corresponding characters. The language is based on nested constructs of the S-expression with the super-characters which employs rather simple structures and appropriate ways as well as to make generalization. In this paper, a modeling environment for the language is presented. Syntax of the language, architecture and interface of the environment are demonstrated.

[yutaka-17:2012] Haruka Hashimoto. Filmification of Methods: Image Processing in *AIDA, University of Aizu, 2012.

Filmification of methods is to represent algorithms in AIDA. *AIDA is a programming language representing algorithms by pictures and moving pictures used as super-character. This paper shows the filmification of methods for image processing. For doing this, necessary components are developed such as pictures used as super-characters which define computational scheme in *AIDA. Also, several template programs are implemented for image processing. As a result of the implementation, several image processing algorithms can be implemented in *AIDA. The user can implement image processing algorithms easily by using clarity pictures and data structures. In *AIDA, beginners who lack programming skills can implement image processing programs without the knowledge of image processing and image data structures. In addition, the user can apply what other image processing algorithms on the basis of this implementation result.