Database Systems Laboratory

Subhash Bhalla
Professor

Yutaka Watanobe
Senior Associate Professor

Wanming Chu
Assistant Professor

Information systems in scientific data archives, transportation, welfare, healthcare 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 e-governance and services use database systems as a basic part for web data resources. This year the laboratory organized an International Symposium with many invited distinguished researchers during 27 Nov. - 29 Nov. 2017. 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 symposium. The delivered lecture provided recent views on 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: a) developing new user interfaces and query languages for skilled and semi- skilled users in health-care; b) developing infrastructures for computing facilities for cloud computing; c) supporting mobile computing applications, d) designing new data models and asynchronous computation models for trans- action and services, and e) Study of Poly-store Data Management Systems. 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 in- formation management. The applications will utilize a shared base of information. Some examples are - business automation, industrial automation, computer-aided design and manufacture, and cartography. Workflow 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 prototypes have been evolved based on research on new easy-to-use search and also new query language interfaces.

Refereed academic journal

[bhalla-307-008-01:2017] Subhash Bhalla Shelly Sachdeva, Shivani Batra. Evolving large scale healthcare applications using open standards. *Health Policy and Technology*, 6(4):410–425, December 2017.

Electronic Health Records (EHRs) are becoming more prevalent in health care. Worldwide exchange of healthcare data demands adherence to semantic interoperable standards to overcome the language and platform barriers. Various healthcare organizations in developing countries such as, India adopt their own independent information systems without adhering to standard guidelines. Thus, this tends to sacrifice interoperability. This affects permanent persistence of longitudinal health records for future reference and research purpose. Current research implements a standard based clinical application to be used for healthcare domain in India. The study has been done for enhancing the data quality through standardization. It aims at providing a generic permanent persistence to track life-long interoperable health records of patients. This is the first effort for exploring its adoption for various regional languages in India. The user interfaces have been generated for various Indian languages for testing on a sample set of archetypes. The clinical application deployed in Hindi language can be easily deployed for other people in Tamil language, while maintaining semantic interoperability. The persistence will also be maintained, with the same meaning (of data) for both the regions. Implementing these standard based healthcare applications helps in reducing the costs while enhancing patient care. Thus, this study aims to build a standard based, and platform independent healthcare application to provide support for interoperability, usability and generic persistence.

[bhalla-307-008-02:2017] Subhash Bhalla Shivani Batra, Shelly Sachdeva. Entity Attribute Value Style Modeling Approach for Archetype Based Data. *Information*, 9(1):1–30, April 2018.

Entity Attribute Value (EAV) storage model is extensively used to manage healthcare data in existing systems, however it lacks search efficiency. This study examines an entity attribute value style modeling approach for standardized Electronic Health Records (EHRs) database. It sustains qualities of EAV (i.e., handling sparseness and frequent schema evolution) and provides better performance for queries in comparison to EAV. It is termed as the Two Dimensional Entity Attribute Value (2D EAV) model. Support for ad-hoc queries is provided through a user interface for better user-interaction. 2D EAV focuses on

how to handle template-centric queries as well as other health query scenarios. 2D EAV is analyzed (in terms of minimum non-null density) to make a judgment about the adoption of 2D EAV over n-ary storage model of RDBMS. The primary aim of current research is to handle sparseness, frequent schema evolution, and efficient query support altogether for standardized EHRs. 2D EAV will benefit data administrators to handle standardized heterogeneous data that demands high search efficiency. It will also benefit both skilled and semi-skilled database users (such as, doctors, nurses, and patients) by providing a global semantic interoperable mechanism of data retrieval.

[yutaka-307-008-01:2017] W. Chen, Y. Yaguchi, K. Naruse, Y. Watanobe, and K. Nakamura. QoS-aware Robotic Streaming Workflow Allocation in Cloud Robotics Systems. *IEEE Transactions on Services Computing*, PP(99):1–1, 2017.

Current solutions of computation offloading for cloud robotics face challenges: 1) traditional approaches do not consider the characteristics of networked cloud robotics (NCR)(e.g., heterogeneity and robotic cooperation); 2) they fail to capture the characteristics of tasks in a robotic streaming workflow (RSW) (e.g., strict latency requirements and different task semantics); and 3) they do not consider quality-of-service (QoS) issues for cloud robotics. In this paper, we address these issues by proposing a QoS-aware RSW allocation algorithm for NCR with joint optimization of latency, energy efficiency, and cost, while considering the characteristics of RSW and NCR. We first propose a novel framework that combines robot individuals, robot clusters, and a remote cloud for computation offloading. We then formulate the joint QoS optimization problem for RSW allocation in NCR while considering latency, energy consumption, and operating cost, and show that the problem is NP-hard. Next, we construct a data flow graph based on the characteristics of RSW and NCR, and transform the RSW allocation problem into a mixed-integer linear programming problem. To obtain an optimal solution in reasonable time, we also develop a heuristic-based algorithm. Experiments demonstrate significant performance gains, with improved QoS and reduced execution times.

Refereed proceedings of an academic conference

[bhalla-307-008-03:2017] Shivani Batra Subhash Bhalla, Shelly Sachdeva. Semantic Interoperability in Electronic Health Record Databases: Standards, Architecture and e-Health Systems. In *Proceedings of 5th International*

Conference on Big Data Analytics, volume 10721 of *Lecture Notes in Computer Science*, pages 235–242. Springer, December 2017.

Information systems have been deployed in different clinics and hospitals to preserve patient data. In order to promote the exchange of data among systems (and organizations), standards are being adopted for data exchange. Further, the clinics and hospitals aim to manage a patient's life-time history of records. A piece of the individual patients medical record can be captured, stored, queried, and shared over a network through enrichment in information technology. Thus, electronic health records (EHRs) are being standardized for incorporating semantic interoperability. In addition, a generic storage structure is required to capture distinguished data requirements of various organizations. The generic structure must be capable of dealing with sparseness and frequent evolution behavior of EHRs. A subsequent step requires that healthcare professionals and patients get to use the EHRs, with the help of technological developments, such as workflow toolkits and new (easy) query languages. The goal is to present an overview of different approaches in understanding some current and challenging concepts in e-health informatics. Successful handling of these challenges will lead to improved quality in healthcare by reducing medical errors, decreasing costs, and enhancing patient care. The report is focused on the following objectives: (1) understanding the role of EHRs Databases; (2) understanding the need for standardization to enhance quality; (3) establishing interoperability in maintaining EHRs; (4) explicating a framework for standardization and interoperability (the openEHR architecture); (5) exploring various data models for managing EHRs; and (6) understanding the difficulties in querying data in EHR and e-health systems.

[bhalla-307-008-04:2017] Subhash Bhalla Mittapally Kumara Swamy, P. Krishna Reddy. Association Rule Based Approach to Improve Diversity of Query Recommendations. In *International Conference on Database and Expert Systems Applications (DEXA 2017)*, volume 10439 of *Lecture Notes in Computer Science*, pages 340–350. Springer, August 2017.

Query recommendation (QR) support search engine to provide alternative queries as a recommendation using similarity-based approaches. In the literature, orthogonal query recommendation (OQR) has been proposed to compute the diversity of QR when the user does not formulate proper queries. The OQR uses dissimilarity measure in QR to recommend completely different queries. In this paper, we propose an approach in QR by extending association rules, diverse patterns, and unbalanced concept hierarchy of search terms. We concep-

tualize association rules based QR, and order the rules based on confidence and diversity. Subsequently, the high ranked rules based on confidence and diversity are provided in QRs. The experimental results on real world AOL click-through dataset show that the diverse QRs improve the performance significantly.

[bhalla-307-008-05:2017] Alexander Vazhenin Ruth Cortez Subhash Bhalla Shashank Shrestha Absalom Shu, Konstantin Markov. Unified User-Interface and Protocol for Managing Heterogeneous Deep Learning Services. In *New Trends in Intelligent Software Methodologies, Tools and Techniques*, volume 297 of *Frontiers in Artificial Intelligence and Applications*, pages 563–575. IOS Press, September 2017.

n the last decade, cheaper and more powerful computations have favored a sufficient surge in research and development of applications in the fields of machine and deep learning. Though often varying in approach, these activities aim mostly at solving similar tasks such as speech synthesis, emotion detection, image recognition, mathematical computations etc. Usually, the typical scenario of using designed algorithms/applications includes inputting data represented in some predefined formats and launching a corresponding program tool that produces expected output data as well as analyzing obtained results. These applications are numerous, and are created by different research teams and laboratories using different techniques and environments including locations, interfaces and procedures to access, query and use them. In a collaborative working environment, developing independent user interfaces for each deep learning application could entail a lot of additional development efforts. In the presented paper, we propose a standardized and flexible interface to reduce design efforts, based on integration of various Deep Learning services (DL services). We also demonstrate a protocol for communication between the user interface and the heterogeneous services. This platform will enable developers of deep learning models to be concerned solely with developing and tuning their models, which can be easily plugged into the central user interface, conveniently exposing their services to users who will have homogeneous central access to a wide-range of DL services, from a unified interface.

[bhalla-307-008-06:2017] Yilang Wu Ruth Cortez Alexander Vazhenin Absalom Shu Subhash Bhalla Shashank Shrestha, Wanming Chu. Workflow Based Query Management System for Astronomical Data Repository. In *New Trends in Intelligent Software Methodologies, Tools and Techniques*, volume 297, pages 719–730. Frontiers in Artificial Intelligence and Applications, IOS Publisher, September 2017.

Astronomy is a data-intensive science. The large amount of data available in the astronomical domain needs query languages to gain valuable astronomical information. The paper has two goals. First, it presents the challenges for domain-specific query language for managing astronomical data. Second, it proposes a solution of managing such large amount of data through incorporating a multi-stage query language with Workflow Management technique. The paper demonstrates a web based query management system able to handle user queries in single or multi-stage.

[bhalla-307-008-07:2017] Naman Jain Shelly Sachdeva Shivani Batra Subhash Bhalla Prateek Jain, Sagar Bhargava. Healthsurance - Mobile App for Standardized Electronic Health Records Database. In *VLDB 2017 Workshop on Data Management and Analytics for Medicine and Healthcare*, Lecture Notes in Computer Science, pages 136–153. Springer, 2017.

With the increasing popularity of Electronic Health Records (EHRs), there arises a need to understand its importance in terms of clinical contexts for a standard based health application. Standards for semantic interoperability propose the use of archetypes for building a health application. A usual practice followed for storing of EHRs is through graphical user interfaces. Generally, user interface is static corresponding to the underlying medical concept, often made manually and are prone to errors. However, evolution in knowledge demands for dynamically generated user interfaces to reduce time, minimize cost and enhance reliability. Current research implements mobile app for standardized Electronic Health Records Database termed as HEALTHSURANCE. The application maintains its dynamic behavior through creation of graphical user interfaces at runtime by gaining knowledge from the artefacts (known as archetypes) available from standard clinical repositories (such as Clinical Knowledge Manager). This provides easy and hassle-free user operability without any need of mobile developer. A standardized format and content helps to uplift the credibility of data and maintains a uniform and specific set of constraints used to evaluate the user's health. A generic centralized database is chosen for data storage to support evolution in clinical knowledge and to handle heterogeneity of EHRs data. Implementing mobile app based on archetype paradigm avoids reimplementation of systems, migrating databases and allows the creation of future-proof systems.

[w-chu-307-008-01:2017] Yamin Li and Wanming Chu. KMS-Cube - A general alternative to hypercubes for reducing the node degree. In *International*

Conference on Computer, Information and Telecommunication Systems, CITS2017, pages 316–320, Dalian, China, July 2017. IEEE.

As the scale of parallel computer systems becomes larger year by year, interconnection networks with low node degree and short diameter are required for achieving high communication performance at low cost. This paper presents a general alternative to hypercubes, named KMS-Cube, that keeps interesting topological properties of the hypercube but has a low node degree. A KMS-Cube, denoted by KMSC(k, m, s), is configured with three parameters k, m, and s. The node address has $m2^(k - s) + k$ bits and the node degree is $k + m$, much less than that of hypercubes. Comparing to other alternatives to hypercubes, KMS-Cube has advantages such as the ease of routing and the flexibility of changing network size. We describe the KMS-Cube's structure and its topological properties, compare it with other networks on the measures of node degree, diameter, and cost ratio, and present an efficient routing algorithm for the KMS-Cube.

[yutaka-307-008-02:2017] T. Yamashita and Y. Watanobe. Consensus Building Algorithm with BFT for Permissioned Blockchain. In *World Congress 2018*, February 2018.

In the blockchain technology, distributed consensus building algorithms are employed so that multiple peers can keep the same states. Generally, the consensus building algorithm in the blockchain must have a feature of Byzantine Fault Tolerance (BFT) which guarantees proper operations even if some peers with Byzantine obstacles are involved in the ledger. Although, for permissioned blockchain, the consensus building algorithms can be applied to the blockchain with a basis of PBFT, another mechanism called OrderingService for constructing a block of transactions is required. Besides, algorithms with BFT which consider all phases from transactions provided by clients to storing data into the blockchain, have not been published. In this paper, a novel algorithm with strict BFT for the blockchain which has a basis of PBFT, is presented.

[yutaka-307-008-03:2017] C.M. Intisar and Y. Watanobe. Fuzzy Rule Mining by Clustering Approach to Estimate the Difficulty of Programming Problems. In *World Congress 2018*, February 2018.

Programming is one of the vital skill for the next generation. Currently there exists many online platforms where programmers compete and solves programming problem. Those platforms are composed of problems with varying degree of difficulty. For expert programmer the difficulty level is not a concern, but it

is very important for novice programmers to approach programming problems based on their experience and level. Thus there rises a need for an expert system which can categorize the programming problems based on their difficulty. In our research we have proposed a knowledge based system which is implemented based on fuzzy rules derivation. These fuzzy rules have been derived from cluster analysis of programming problems. Later, inference system has been build based on these rules and knowledge to estimate the difficulty of the programming problems.

Unrefeered proceedings of an academic conference

[yutaka-307-008-04:2017] D. Yoshino, Y. Watanobe, Y. Yaguchi, K. Nakamura, J. Ogawa, and K. Naruse. Proposal of MQTT and MQTT-SN Communication Interfaces on RT Middleware for IoR System Construction. In *The 18th Meeting of SICE System Integration Department, SI2018*, 2017.

[yutaka-307-008-05:2017] D. Yoshino, Y. Watanobe, Y. Yaguchi, K. Nakamura, and K. Naruse. Application possibility of OpenRTM-aist-based integrated robot systems using CORBA interfaces and brokered Pub/Sub messaging interfaces. In *2017 JSME Conference on Robotics and Mechatronics, ROBOMECH2017*, 2017.

[yutaka-307-008-06:2017] K. Amma, Y. Yaguchi, Y. Watanobe, and K. Naruse. Constructing Cloud base RTM and automatic deploy to Raspberry Pi. In *2017 JSME Conference on Robotics and Mechatronics, ROBOMECH2017*, 2017.

Writing a textbook or technical book

[bhalla-307-008-08:2017] Sharma Chakravarthy Subhash Bhalla P. Krishna Reddy, Ashish Sureka. *Big Data Analytics - 5th International Conference, BDA 2017, Hyderabad, India, December 12-15, 2017, Proceedings*, volume 10721. Springer, December 2017.

[bhalla-307-008-09:2017] Bhateja V. Chandavale A. Hiwale A.S. Satapathy S.C. Bhalla, S. *Intelligent Computing and Information and Communication*. Advances in Intelligent Systems and Computing. Springer, August 2017.

Summary of Achievement

The volume presents high quality research papers presented at Second International Conference on Information and Communication Technology for Intelligent Systems (ICICC 2017). The conference was held during 2 - 4 August 2017, Pune, India and organized communally by Dr. Vishwanath Karad MIT World Peace University, Pune, India at MIT College of Engineering, Pune and supported by All India Council for Technical Education (AICTE) and Council of Scientific and Industrial Research (CSIR). The volume contains research papers focused on ICT for intelligent computation, communications and audio, and video data processing.

Research grants from scientific research funds and public organizations

[bhalla-307-008-10:2017] Subhash Bhalla. SCAT research subsidies 2017-2018, 2017.

Academic society activities

[yutaka-307-008-07:2017] Y. Watanobe, August 2017.

Program Committee, The International Conference on Big data, IoT, and Cloud computing.

[yutaka-307-008-08:2017] Y. Watanobe, December 2017.

Program Committee, The 9th International Conference on Computer Science and its Applications

[yutaka-307-008-09:2017] Y. Watanobe, December 2017.

Organizing Committee Member, ACM-ICPC Asia Regional 2018 Tsukuba

Advisor for undergraduate research and graduate research

[w-chu-307-008-02:2017] Masashi Baba. Graduation Thesis: Improvement of Advanced Bus Route Search System using Google Maps API, University of Aizu, 2017.

Thesis Advisor: Wanming Chu

[yutaka-307-008-10:2017] Takumi Yamashita. Graduation Thesis: Consensus Building Algorithm with BFT for Permissioned Blockchain, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

[yutaka-307-008-11:2017] Motohiko Abe. Graduation Thesis: Rose: New Byzantine Consensus on Hashgraph based Data Structure, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

[yutaka-307-008-12:2017] Keigo Ebihara. Graduation Thesis: Evaluation of Language Network System, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

[yutaka-307-008-13:2017] Yunosuke Teshima. Graduation Thesis: Bug Detection based on Deep Learning and Solution Source Codes, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

[yutaka-307-008-14:2017] Ryoya Komatsu. Graduation Thesis: Online Text Editor with Logical Error Correction, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

[yutaka-307-008-15:2017] Kazuya Watanabe. Master Thesis: Grading Algorithm using Difficulty Level Estimation of Problem Sets, University of Aizu, 2018.

Thesis Advisor: Y. Watanobe

Others

[yutaka-307-008-16:2017] Y. Watanobe, Feburary 2018.

Programming Challenge on New Aizu Online Judge

Contributions related to syllabus preparation

[yutaka-307-008-17:2017] A undergraduate school course syllabus constructed: [F01] Algorithms and Data Structures

[yutaka-307-008-18:2017] A undergraduate school course syllabus constructed: [IE03] Integrated Exercise for Software I

[yutaka-307-008-19:2017] A undergraduate school course syllabus constructed: [OT04] Courses for the Information Technology Examination

[yutaka-307-008-20:2017] A graduate school course syllabus constructed: [CSC11F] Advanced Data Structures and Algorithms

[yutaka-307-008-21:2017] A graduate school course syllabus constructed: [SEC04A] Programming Strategies and Software Development Tools

Preparation of course examination to measure comprehension

[yutaka-307-008-22:2017] Problem Setter for Entrance Exam Questions of Mathematics

Advisor of a student club or circle

[yutaka-307-008-23:2017] Circle Advisor: Competitive Programming Club (ICPC), ACM-ICPC World Finals 2017, ACM-ICPC Asia Regional 2017 Tsukuba, ACM-ICPC Asia Regional 2017 Hualien, ACM-ICPC Asia Regional 2017 Daejeon, ACM-ICPC Asia Regional 2017 Ho Chi Minh

Other significant contribution toward university planning, management, or administration

[yutaka-307-008-24:2017] A member of Judge for Programming Division of PC Koshien

[yutaka-307-008-25:2017] A member of entrance examination committee

Contributions related to regional education

[yutaka-307-008-26:2017] Special Lecture: Programming Education for Elementary Schools, Koriyama Central Public Hall, 2017, July

Contribution toward education for employees of regional industries

[yutaka-307-008-27:2017] Lecture for RT Middleware in Minamisoma-City, Minamisoma, 2017, August

[yutaka-307-008-28:2017] Software for Robot, Koriyama, 2017, August

[yutaka-307-008-29:2017] Fukushima Human Resource Development Curriculum, Koriyama, 2018, Feburary

Did you participate in Public Lectures, and/or Open Campus? (Yes or No) If yes, please describe what you did.

[yutaka-307-008-30:2017] Trial Lesson of Programming, Gakuho High School, 2017, July

[yutaka-307-008-31:2017] Sports Programming and Online Judge System, Lecture for Computer Science Summer Camp, 2017, August

[yutaka-307-008-32:2017] Introduction to Programming, Kitakata High School, 2017, October

[yutaka-307-008-33:2017] Introduction to Programming, Yamagata Institute High School, 2017, November

[yutaka-307-008-34:2017]

[yutaka-307-008-35:2017]

[yutaka-307-008-36:2017]

Research achievement that can be used for University-Industry collaboration and its characteristics.(for UBIC's information)

[yutaka-307-008-37:2017] Online Judge System (Aizu Online Judge): https://onlinejudge.u-aizu.ac.jp/

[yutaka-307-008-38:2017] Visual Programming Language and Environment (*AIDA): http://aida.u-aizu.ac.jp/