

第13回

The 15th International Conference on Database Systems for Advanced Applications (DASFAA2010)

2010年4月1日～4日
筑波大学大学会館(茨城県つくば市)

北川博之^{*1} 石川佳治^{*2} 天笠俊之^{*1}

^{*1} 筑波大学大学院システム情報工学研究科 ^{*2} 名古屋大学情報基盤センター

DASFAA について

第15回 International Conference on Database Systems for Advanced Applications (DASFAA2010) が、2010年4月1日～4日に茨城県つくば市の筑波大学大学会館で開催された。DASFAA は、日本のデータベースコミュニティ等が中心となって立ち上げた、データベース、データマイニング、Web などの大規模データ利用に関する国際会議である。開催地は、東アジア、オーストラリアなどの環太平洋地域を中心に回っているが、国際的に認知される国際会議として成長を遂げた。第1回は1989年に韓国ソウルで開催され、以降日本では東京(第2回, 1991年)、京都(第8回, 2003年)に引き続き7年ぶり3回目の開催となった。データベース関係の国際会議では、ICDE2005(東京)、MDM2006(奈良)以来、4年ぶりである。

本コラムでは、北川(会議委員長)、石川(プログラム共同委員長)、天笠(ローカルアレンジメント委員長)が、主催者側の立場から DASFAA2010 の概要について報告する。

開催までの経緯と準備

DASFAA2010 のつくばでの開催が正式に決定されたのは、DASFAA2007 における運営委員会(Steering Committee) でのことある。それ以来、



写真1 大学会館(総合交流会館)

準備に約3年を費した。

当初は、開催時期は4月中旬、会場はつくば国際会議場とする方向で準備を開始した。しかし、2008年9月のリーマンショック以降、経済状況が激変し、参加者の減少や収支の悪化が懸念されるようになった。これらの状況を総合的に判断して、2009年の早い時期に、会場を筑波大学キャンパス内の大学会館に変更することとし、会期も桜の開花時期等を考慮して4月1日～4日とすることとした(写真1)。

ただし、会議初日が年度初めに当たることを考慮し、初日は昼から会議を開始することとして国内の移動に配慮した。一方、特に欧米からの参加者に関して、この週末がイースター(復活祭)と重なってしまうという点が懸念された。イースターは、日本



写真2 会場付近の桜

人にはあまり馴染みがないため、当事者に指摘されないと気づくのはなかなか難しい問題である。

結果的には、当初予定していた4月中旬には、季節外れの低気温と降雪、さらにはアイスランドの火山噴火による航空網の混乱などがあった。また、イースターに関してはほとんど影響がなかったように思われる。また、若干咲き始めの時期ではあったが、会期中に桜の花も楽しめたということで、日程を変更したのは正解であった(写真2)。

会議の約1年前からは本格的かつ組織的な準備を進めるために「準備委員会」を定期的で開催した。初回を2009年3月に開催し、以降、原則月1回、計11回開催した。会場には都内の大学、研究所の会議室をお借りし、関東圏以外の関係者とは、テレビ会議システム(Polycom)とSkypeを併用して接続した。準備委員会は毎回夜の時間帯に開催したため、参加者には簡単な夕食を用意し、食事をしながらときには10時近くまで議論が白熱することもあった。

論文投稿

国際会議の運営で最も重要なことの1つは、いかに多くの良質の論文投稿を得るかという点である。このため、論文投稿スケジュールの設定には細心の注意を払った。関連分野の主要国際会議のオーガナイザには事前に連絡を取り、投稿時期が重ならないようにスケジュールリングを心掛けた。さらに、各種

Research Session	
1	P2P-Based Technologies
2	Data Mining Technologies
3	XML Search and Matching
4	Graphs
5	Spatial Databases
6	XML Technologies
7	Time Series and Streams
8	Advanced Data Mining
9	Query Processing
10	Web
11	Sensor Networks and Communications
12	Information Management
13	Communities and Web Graphs
14	Trajectories and Moving Objects
15	Skyline Queries
16	Privacy and Security
17	Data Streams
18	Similarity Search and Event Processing
19	Storage and Advanced Topics
Industrial Session	
1	Cloud/SaaS
2	Query Processing/Data Mining

表-1 セッション一覧

チャンネルを通じて積極的な広報活動を行った結果、リサーチセッションについては、25の国と地域から237件の投稿があった。これは過去のDASFAAに比べても高い水準の投稿数となった。

投稿論文は、99名のプログラム委員による査読の審査を行い、55件のフルペーパー、16件のショートペーパーを採択した。フルペーパーの採択率は23%であり、採択論文の質は高いレベルにあると言える。日本からの投稿については、29件の投稿中3件採択で採択率は1割程度にとどまり、やや厳しい結果となった。日本のデータベースコミュニティの今後の奮起が強く望まれる。採択論文は、19のリサーチセッションに取りまとめた(表-1)。以下では、データベース、データ工学分野における研究の動向とDASFAA2010での状況について紹介する。

データベース、データ工学分野における 研究動向と DASFAA2010 での状況

データベース、データ工学分野は、その名が示すとおり、データベースやデータベースシステム(DBMS)そのものに関する研究、あるいはDBMSおよび関連技術を核とした大規模データの利活用を目的とした研究分野である。データベースやDBMSそのものを対象としたコアの技術としては、データベース設計、索引、問合せ最適化、トランザクション処理などがあり、またデータベースからのデータマイニングなども分野に含まれる。

しかしながら、情報源の多様化、DBMS技術の成熟とコモディティ化に伴い、研究の最前線はデータベースシステムそのものに関するトピックよりは、Webやストリームなどの新たな情報源からのデータマイニングや情報抽出、あるいはP2Pやクラウド等の新たなコンピューティングプラットフォームを用いた大規模データ処理にシフトしてきている。DASFAA2010も投稿の傾向は同様であった。以下に、今回目立ったトピックをいくつか紹介する。

[データの曖昧性や信頼性等を考慮した問合せ処理やマイニング]

従来のデータベースでは、確定した事実のみがデータベースに格納されることを前提としていた。しかしながら、応用によってはデータや問合せの結果が曖昧性を含む場合が少なくない。たとえば移動体の位置情報を考えると、GPSやセンサの性質上必ず測定誤差が含まれる。このため問合せ処理は位置の曖昧性を考慮する必要がある、問合せ結果には確信度が付与されることとなる。このようなデータの不確実性や信頼性を考慮した問合せ処理やマイニングは近年盛んに研究されており、多くのセッションでデータの曖昧性や信頼性等を考慮した問合せ処理やマイニングに関する発表が見られた。

また、大量のレコードに対して問合せを行う場合、利用者は必ずしもすべての問合せ結果に興味があるとは限らない。この場合、利用者の情報要求になる

べく沿うような一部(上位k件)のレコードを確信度などのランキングに従い優先的に返却することが重要である。このような処理をTop-k問合せ¹⁾といい、これに関しても多くのセッションで発表があった。

[ストリーム処理]

ストリームデータとは、情報源から連続的に到着するデータのことであり、近年注目されている種類のデータである。典型的な情報源としてはセンサネットワークなどが容易に想像できるが、意外な例ではWebサイトにおける利用者の閲覧行動といったコンテキストでも用いられる。この場合、利用者の閲覧行動(クリック)が時系列的にクリックストリームとして収集され、問合せやマイニングの対象となる。

今回は、"Research 7: Time Series and Streams", "Research 17: Data Streams"の2セッションがストリーム処理に関係しており、それ以外のセッションでもいくつかの関連する発表が見られた。研究の内容としては、ストリームデータを対象とした各種マイニングアルゴリズムの適用や、未加工の(生の)センサデータからなるストリームデータから、人間の行動などのより高次のイベントを検出するイベント検出処理などが挙げられる。

[スカイライン問合せ]

スカイライン問合せ²⁾とは、複数の数値属性を持つレコード(オブジェクト)群から特徴的なレコードを抽出する処理である。ここで特徴的とは、他のレコードからドミネイト(dominate, 支配)されないことを指す。具体的には、レコードAがすべての属性においてBより優位であるときに、AはBをドミネイトするという。駅からの距離と家賃の2属性を含む不動産物件データベースを考えると、この処理によって距離か家賃の少なくとも一方は、他のレコードに対して優位な(近い、あるいは安い)レコードが抽出され、両方とも負けている(良いところがない)ようなレコードは排除される。このような性質を満たすレコードを、大量(数万~)レコードか

らなるデータベースから高速に抽出することが目的となる。

今回は, "Research Session 15 : Skyline Queries" でスカイライン問合せのセッションがあり, グラフデータや高次元データへの適用など先端的なトピックが議論されている. そのほかにも, "Research Session 1 : P2P-Based Technologies" では, P2P 環境でのスカイライン問合せが, "Research Session 7 : Time Series and Streams" では, ストリームデータに対するスカイライン問合せが議論されている.

そのほか, DASFAA2010 で特徴的だったと感じたのは, XML (Extensible Markup Language) に関する投稿が多かった点である. XML は, 1998 年の W3C 勧告以来標準のデータフォーマットとして急速に普及している. データベース分野でも, それに伴い過去十数年で XML データベースに関する研究がかなり活発に行われた. ただ, この数年その傾向は治まり, 特にトップカンファレンスでは XML 関係の論文は投稿, 採択とも減少傾向にある. これに対して, DASFAA での投稿は引き続き堅調であり, 両者の差は興味深い.

インダストリアルセッション, デモセッションについても各委員長の主導のもと積極的な投稿の呼びかけを行った. インダストリアルセッションでは, 企業における製品やプロジェクト, アプリケーション等に関する論文を募集し, 15 件の投稿中 6 件を採択して, 2 つのセッションにとりまとめた. デモセッションも国内外から多くの投稿があり, 33 件中 22 件を採択した.

参加者について

経済危機による参加人数への影響を懸念していたが, 結果として 300 名を超える 315 名の参加者を得ることができた. 過去の DASFAA の中では, 第 2 回 (東京) に次ぐ多数の参加者数であり, 過去十数年の中では記録的な参加人数となった. 国別の参加者数の内訳を図 -1 に示す. 半数弱が海外からの参加であり, アジアはもちろんのこと, 欧米やアフリ

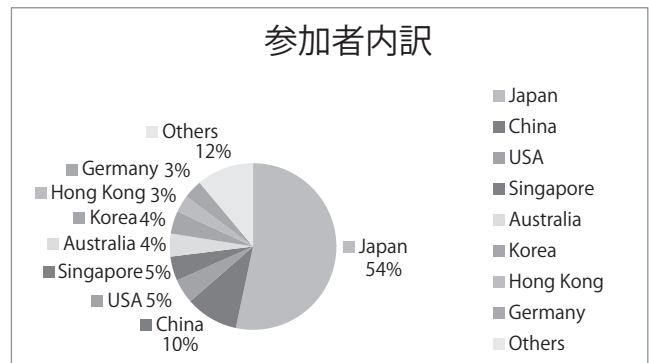


図 -1 参加者内訳

カ大陸など日本を含めて 30 カ国からの参加を得た. 特に, 中国系の参加者の割合は目立っており, 中国本土, 香港だけではなく, 欧米やオーストラリアからの参加も, かなりの数が中国系の研究者であった. 他の分野でも同様の傾向にあると思われるが, この分野における中国勢の躍進を改めて印象付けられた.

プログラムハイライト

DASFAA2010 は, プログラムの内容面でも参加者から高い評価をいただいた. ここではそのハイライトを紹介する.

[キーノート]

データベース分野では世界的にきわめて著名な 2 名の研究者から基調講演をいただいた. Gerhard Weikum 博士 (独マックスプランク研究所) からは, "Knowledge on the Web : Robust and Scalable Harvesting of Entity-Relationship Facts" と題して, Web からの知識獲得に関する取り組みについて講演いただいた (写真 3). Raghuram Ramakrishnan 博士 (米ヤフー研究所) からは, "Cloud Data Management @ Yahoo!" と題したヤフー研究所におけるクラウド環境を用いた大規模データ管理の取り組みについて紹介いただいた (写真 4). 前者が比較的柔らかめのトピック, 後者がデータベースコアに関係する比較的硬めのトピックということで, バランスも良かったのではないと思われる.



写真3 Weikum 博士によるキーノート



写真4 Ramakrishnan 博士によるキーノート

[インダストリアル招待講演]

萩原正義氏（マイクロソフト日本法人）から、"Development Procedure of the Cloud-based Applications" と題したマイクロソフトにおけるクラウドの取り組みに関してインダストリアルセッションの中で講演いただいた。

[チュートリアル]

チュートリアルとしては、"Introduction to Social Computing", "Mining Moving Object, Trajectory and Traffic Data", "Querying Large Graph Databases" の3件を設定した。いずれのチュートリアルも近年特に関心の高まっているトピックに関するものであり、講師もそれぞれ世界的に有名な研究者である。案の定、いずれのチュートリアルも多く立ち見が出る盛況ぶりで、参加者にもきわめて好評であった。

[デモ]

2日目の午後2セッションにデモセッションをアレンジした。デモセッションの開始に先立ち、ランチの時間帯を利用して全デモ発表者による1分間プレゼンテーション(1 minute madness)を行った。デモ自体も非常に盛り上がり活発な議論が交換された(写真5)。



写真5 デモの様子

[表彰]

DASFAA2010 で表彰された論文の一覧を表-2に示す。

ベストペーパー表彰委員会により、リサーチセッションの採択論文の中から、Best Paper, Best Student Paper, ならびにそれらの次点論文が選ばれ、3日目の表彰式で表彰を行った。

また、デモ発表についても、デモ表彰委員会による優秀デモの選定が行われ、ベストデモの表彰を行った。

表彰式では、これに加えて、DASFAA 10+ Years Best Paper Award の表彰も行われた。これは、10年(以上)前のDASFAAで発表された論文の中から、その後の10年で最もインパクトのあった論文を選んで表彰するものであり、今回はDASFAA1995で発表された"Continuous Media Sharing in

Best Paper Award
Detecting Leaders from Correlated Time Series <i>Di Wu, Yiping Ke, Jeffrey Xu Yu, Philip S. Yu, Lei Chen</i>
Best Paper Award Runner-up
Active Duplicate Detection <i>Ke Deng, Liwei Wang, Xiaofang Zhou, Shazia Sadiq, Gabriel Pui Cheong Fung</i>
Best Student Paper Award
Mining Outliers with Ensemble of Heterogeneous Detectors on Random Subspaces <i>Hoang Vu Nguyen, Hock Hee Ang, Vivekanand Gopalkrishnan</i>
Best Student Paper Award Runner-up
k-Selection Query over Uncertain Data <i>Xingjie Liu, Mao Ye, Jianliang Xu, Yuan Tian, Wang-Chien Lee</i>
DASFAA 10+ Years Best Paper Award
Continuous Media Sharing in Multimedia Database Systems, DASFAA 1995 <i>Mohan Kamath, Krithi Ramamritham, Donald F. Towsley</i>
Best Demo Award
MediaMatrix : A Video Stream Retrieval System with Mechanisms for Mining Contexts of Query Examples <i>Shuichi Kurabayashi, Yasushi Kiyoki</i>
Excellent Demo Award
SERPWatcher : A SERP Mining Tool as a Novel Social Survey Method in Sociology <i>Yoshifumi Masunaga, Naoko Oyama, Chiemi Watanabe, Kazunari Ito, Kaoru Tachi, Yoichi Miyama</i>
Fuzzy Keyword Search on Spatial Data <i>Sattam Alsubaiee, Chen Li</i>

表-2 表彰論文一覧

Multimedia Database Systems" が選ばれた。第1著者の Mohan Kamath 博士が受賞後に、受賞記念講演を行った。

[パネル]

"Future Directions of Innovative Integration between Multimedia Information Services and Ubiquitous Computing" と題したパネルディスカッションが5名のパネリストにより行われ、マルチメディア情報サービスとユビキタスコンピューティングの統合の将来像について議論を交わした。最終的には、情報の教育論にまで話がおよび、大変活発なパネルセッションとなった(写真6)。



写真6 パネルセッション



写真7 バンケットでの三味線のアトラクション

ソーシャルイベント

DASFAA2010 では、レセプション(4月1日、大学会館レストラン・プラザ)とバンケット(4月2日、山水亭)の2つのソーシャルイベントをアレンジした。バンケットでは、アトラクションとして琴と三味線の演奏を準備した(写真7)。また桜の生け込みや寿司職人による寿司のサービスなどが参加者に好評であった。

バンケットでは、DASFAA Outstanding Contribution Award の表彰が行われた。これは、DASFAA の発展に長年貢献のあった功労者を表彰するもので、今回は増永良文教授(青山学院大学)と Tok Wang Ling 教授(シンガポール国立大)が受賞さ

The 15th International Conference on Database Systems for Advanced Applications (DASFAA2010)

1st International Workshop on Graph Data Management : Techniques and Applications (GDM2010)
DASFAA2010 Workshop on Large-scale Image and Video Archives (LIVA2010)
2nd International Workshop on Benchmarking of Database Management Systems and Data-Oriented Web Technologies (BenchmarX'10)
3rd International Workshop on Managing Data Quality in Collaborative Information Systems (MCIS2010)
Workshop on Social Networks and Social Media Mining on the Web (SNSMW)
Data Intensive eScience Workshop (DIEW2010)

表-3 ワークショップ一覧

れた。増永教授は、DASFAA の立ち上げ当初から現在に至るまで DASFAA の運営と発展に貢献されたことが評価され、今回の受賞となった。

ワークショップ

最終日（4月4日）はワークショップデーであり、合計6つの併設ワークショップを開催した（表-3）。トピックはソーシャル、ユビキタス、グラフ、XML、eサイエンスなど多岐にわたり、最終日にもかかわらず100名を超える参加者があり、大変活気のある最終日とすることができた。

おわりに

以上、簡単に DASFAA2010 の経緯と概要を報告させていただいた。会期中は、一部雨天や強風があったものの、天候に恵まれ、中ごろには桜の花も楽しむことができた。多くの方から、大変よくアレンジされた会議であったとねぎらいの言葉をいただき、

主催者一同、大いに報われる思いである。これも、ひとえに、種々ご助言を下された先輩諸氏、委員として精力的にご尽力いただいた多くの皆様、ご支援いただいたさまざまな組織等、関係各位のお力添えの賜物であり、心より御礼を申し上げたい。

準備にほぼ3年を費やしたものの、会議が始まってみればあっという間の4日間であった。世界中から関連分野の研究者が一堂に集い、皆が会議を楽しみ、笑顔で帰る様子を見届けるのは、会議を運営することの醍醐味である。実際、規模の大きな国際会議を日本で開催するには、莫大なエネルギーを投入しなければならないが、そのコミュニティにおける日本のプレゼンスを向上させ、海外研究者との連携を強めることができる点できわめて意義深いものであることを改めて実感した。貴重な機会をいただいたことに感謝したい。

参考文献

- 1) Fagin, R., Kumar, R. and Sivakumar, D. : Comparing Top k Lists, SIAM J. Discrete Math, Vol.17, No.1, pp.134-160 (2003).
 - 2) Börzsönyi, S., Kossmann, D. and Stocker, K. : The Skyline Operator, Proc. ICDE 2001, pp.421-430 (2001).
- (平成22年4月26日受付)

北川博之（正会員） | kitagawa@cs.tsukuba.ac.jp

1980年東京大学理学系研究科修了。現在、筑波大学大学院システム情報工学研究科教授、兼計算科学研究センター教授。理学博士。研究分野は、データベース、データ工学。電子情報通信学会、本会フェロー。

石川佳治（正会員） | ishikawa@itc.nagoya-u.ac.jp

1994年筑波大学工学研究科単位取得退学。奈良先端科学技術大学院大学、筑波大学を経て、2006年より名古屋大学情報基盤センター教授。博士（工学）。データベースの研究に従事。

天笠俊之（正会員） | amagasa@cs.tsukuba.ac.jp

1999年群馬大学大学院工学研究科修了。奈良先端科学技術大学院大学助手、筑波大学大学院システム情報工学研究科講師を経て、2009年より同准教授。計算科学研究センター准教授を兼任。博士（工学）。データベースの研究に従事。