



## 会議レポート

# 22<sup>nd</sup> International Conference on Extending Database Technology (EDBT) 参加報告

### EDBT 2019 概要

本稿は2019年3月26日～29日にポルトガルのリスボンにて開催されたデータベースの国際会議、22<sup>nd</sup> International Conference on Extending Database Technology (EDBT) に関する参加報告である。なお EDBT は SIGMOD/PODS の同時開催に学び、2009 年から International Conference on Database Theory (ICDT) と呼ばれるデータベース理論の会議と同時開催を行っているが、本稿では EDBT に焦点を絞り、お伝えする。

EDBT はデータベースの研究トピックを幅広く扱う学会であり、ストレージ関連からインデックスやクエリ処理、ストリームデータ処理や機械学習まで多岐に渡る。同時開催の ICDT と比べると実践的な研究が多い印象で、新たなアルゴリズムの提案や既存手法の改良はもちろんであるが、それらを用いて作った処理機構、あるいはデータベースそのものの研究発表も見られた。投稿形式は Full Paper, Industrial & Application Paper, Short Paper, Demo の 4 種類があり、前者 2 種はオーラルセッションにて 30 分のプレゼン+質疑による発表、後者 2 種はポスター発表となる。2019 年は、Full Paper が 157 件中 36 件、Industrial & Application Paper が 24 件中 8 件、Short Paper が 122 件中 28 件、Demo が 42 件中 21 件採択された。また、そのほかに ICDT と EDBT が合同で行うワークショップが 5 件、チュートリアルが 3 件あり、キーノートは 4 件であった。全体的に採択率は 25% 程度ということで、近年の EDBT と変わらぬ推移を見せており、SIGMOD、VLDB、ICDE に次ぐ国際会議として引き続き存在感を示している。

なお EDBT の特殊な試みとして、2014 年から、過去 10 年以内の最もインパクトのあった研究を表彰する場を設けており、今回はスケーラブルなストレージエンジンの基礎となった 2009 年の論文 “Shore-MT : A Scalable Storage Manager for the Multicore Era” が受賞し、講演が行われた。

EDBT の会場はテークニコ高等学校 (図-1) の中にあ

る IST Congress Center で、ホール 1 つと講義室らしき部屋のいくつかを使って会議は進行した。1 つのセッションは 1 時間半で、最大 4 並列で進む。4 並列のうち Full Paper の発表セッションは最大で 2 つとなっており、残りのセッションではチュートリアルと Industrial & Application Paper 発表が実施される形である (初日と 2 日目は ICDT も開催されているため、1 つは ICDT のセッションであった)。筆者が聴講したセッションはどれも数十人の参加者がおり、参加者の多さと活気が感じられた。セッションの合間には 30 分ほどの休憩がとられ、軽食をとりながら参加者たちが積極的に議論を交わし、親交を深めていた。EDBT のポスターセッションは 27 日と 28 日における最後のセッションに行われ、吹き抜けの広い空間にポスターとデモが立ち並び、大変な賑わいを見せた。こちらも軽食が用意され、ポスター発表者だけでなく参加者同士の交流も盛んで、小さな懇親会といった様相でもあった。また、27 日夜には会場近くでのレセプションがあり、28 日夜にはバスで移動してのバンケットで夜景と歌を楽しみながら食事をとった。

本稿では、以降、キーノート、オーラルセッション、チュートリアルについてそれぞれの内容や傾向と、得た印象について述べる。

### キーノート

現在の流行りとも言える機械学習については EDBT においても多くの注目を集めており、2 つのキーノートが機械学習に関連した発表を行った。データベースと機械学習がどのようにかかわっていくかについて、“Learning Models over Relational Databases” というキーノートでは、データベース研究者ならではの発想で機械学習を扱う方法を提案した (図-2)。すなわち、データベース内に入ったデータを用いて機械学習を行う場合においては、大量のデータベース内のデータをいったん 1 つの表形式にまとめ、エクスポートする部分が非常に冗長な処理となっており、データベース内で SQL クエリのバッチとして



図-1 テークニコ高等学校

機械学習を実行することで学習を大幅に高速化できるというものである。有用な場面は限定されるが、非常に興味深く感じられた。もう1つは“The Power of Relational Learning”と題しており、機械学習に代わる宣言的な関係表現による学習（Relational Learning）についての講演であった。ほかにも2日目に行われたパネルセッションでは、データベースの学会でありながら機械学習が論点として大きく取り上げられ、会場からも多くの質問が飛び出し、参加者の関心の高さをうかがうことができた。

残り2つのキーノートはそれぞれ“Fact Checking”と呼ばれるフェイクニュースなどの事実を判定する研究と、検索クエリに主観的な条件（たとえば、「部屋が綺麗」など）を取り入れて検索を行うための“Subjective Database”を自然言語処理分野の技術を用いて構築する研究であった。先の機械学習もそうであるが、データベースシステムの中核だけでなく、さまざまな周辺分野の考えや手法を取り入れた研究が注目されていることが分かる。

## オーラルセッション

オーラルセッションは2並列で進む Full Paper セッションと Industrial & Application Paper セッションがあるが、筆者は主に Full Paper のセッションを中心に聴講した。プログラムで目を引いたのは、グラフデータベースおよびセマンティック Web に関する研究である。今回の EDBT ではこのセッションは2回設けられており、採録数そのものが多い。このほかに2回セッションが設けられているのはデータマイニングのセッションのみである。グラフデータベースやセマンティック Web における膨大で困難な処理に対し、実行計画の効率化から新たなサンプリングによる計算手法まで、多方面からアプローチが行われているようだ。

他方では、今回最優秀論文に選ばれた“Efficient Window Aggregation with General Stream Slicing”のように、ストリーム処理に関する研究も散見された。ストリーム処理はバッチ処理と違い、時間的に連続して到着

するデータに対してさまざまな処理を行う。そのため、到着順の混乱や複数のウィンドウ処理など多くの複雑さを抱えている。EDBT においてもそのストリーム処理特有の難しさに関して解決策が提案されていた。特に最優秀論文はストリーム処理の中でも集計処理に着目し効率的な処理方法を提案するだけでなく、まだ見ぬ集計関数にも後から対応可能なフレームワークとして構築しており、非常に実用的な研究となっている。

従来からのストレージやインデックス、クエリ最適化といったトピックは、キーノートやチュートリアルではあまり注目されている印象ではなかったが、オーラルセッションでの研究発表は盛んに行われていた。たとえば FPGA を活かし、従来のオンライントランザクション処理のボトルネックであったインデックスの処理を効率化する研究や、多次元ワークロードの特にレンジクエリを対象とし、適したインデックス構造を作ることで効率化する研究などがある。ただ、筆者の観測範囲内では NoSQL に関する研究が少なく、リレーショナルデータベース以外のデータベースでは先に述べたグラフデータベースが多数を占めているのが印象的であった。

## チュートリアル

EDBT 2019 のチュートリアルは3セッションあり、それぞれ RDF グラフ、JSON スキーマ<sup>☆1</sup>、ソーシャルネットワークグラフの Influence Maximization についてであった。やはりグラフに関するチュートリアルが多数を占め、グラフデータベースの注目度を察することができた。

筆者は特に分野が近かったために、JSON スキーマのチュートリアルを聴講した。チュートリアルは JSON スキーマを表現する方法からスキーマを用いる処理機構まで網羅的に解説した。聴講者は20人程度おり、質疑も活発に行われていた。JSON に関しては EDBT 2019 のオーラルセッションではほとんど話題にされていないが、筆者も含め参加者の中には注目している人がいることがうかがえた。

## EDBT 2020

2020年の EDBT はデンマークのコペンハーゲンにて、2020年3月30日から4月2日までの開催予定である。次回も ICDT と同時開催となっている。EDBT 2020 の Call for paper はまだ出ていないが、この原稿が世に出るころにはきっと公開されていることであろう。年度末の忙しい時期ではあるが、ぜひ読者諸兄も投稿、あるいは参加してみたいかがだろうか。

(加藤千裕 / (株) 富士通研究所)



図-2 キーノートセッションの様子

☆1 JSON は半構造化データであり、事前にスキーマを定めるテーブルとは違い、1つ1つのデータが柔軟なスキーマを取ることが可能だ。しかし実用上はスキーマを知る、定めることでより広範な処理が可能になるため、JSON スキーマを扱うことは重要である。