

特集 「Linked Data とセマンティック技術」

Linked Data とセマンティック技術の海外動向

Recent Trends in Linked Data and Semantic Technology

川村 隆浩
Takahiro Kawamura
科学技術振興機構情報企画部情報分析室
Japan Science and Technology Agency.
takahiro.kawamura@jst.go.jp

森田 武史
Takeshi Morita
慶應義塾大学理工学部
Faculty of Science and Technology, Keio University.
t_morita@ae.keio.ac.jp

福田 直樹
Naoki Fukuta
静岡大学大学院情報学領域
College of Informatics, Shizuoka University.
fukuta@cs.inf.shizuoka.ac.jp

Keywords: semantics, data, ontology, SPARQL, query.

1. はじめに

Tim B. Lee によってセマンティック Web [Lee 01] が提唱されて、今年で 17 年になる。その間、産業界、公共サービス、または科学分野においてセマンティック技術の広範な適用が進められてきた。特に、Linked Data は政府の透明性を高める施策として、または生命科学や企業間のデータ統合の助けとして、博物館や図書館のデータ公開・統合手段として至るところに普及している。また、産業界では Google や Yahoo!, Microsoft, Facebook のような大企業がそれぞれ独自の Knowledge Graph を構築し、意味的な検索やスマートなデータ処理、配信を実現している。さらに、Google, Microsoft (Bing), Yahoo! による schema.org の活動 (検索結果に詳細情報を反映するために必要な構造化データに関するフォーマットの標準化) は、企業におけるセマンティック Web への関心とサポートを高め、IBM や Siemens, BestBuy, Walmart のような企業はセマンティック Web に関する国際会議の常連となっている。2014 年の ISWC (International Semantic Web Conference) のスポンサー企業は 32 社に上り、Yahoo! Labs, 富士通はプラチナスポンサー、Google, IBM Research, Microsoft Research などがゴールドスポンサーとして名を連ねている。

現在、セマンティック技術に関する国際会議は数多く開催されているが、本稿では本分野で最高峰の ISWC の動向を中心に紹介し、最後にアジア・太平洋地区で日本人コミュニティの貢献が活発な JIST (Joint International Semantic Technology Conference) について紹介する。なお、2016 年 10 月 16 ~ 20 日に兵庫県

神戸市にて ISWC 2016 の開催が決定している。

2. ISWC 2014 参加報告

2.1 全体概要

ISWC 2014^{*1} は 2014 年 10 月 19 ~ 23 日の間、イタリア北部の湖畔の町 Riva del Garda で開催された。ISWC は本分野におけるトップカンファレンスであり、13 回目となる今回は 42 か国から 630 名の参加があり、約 300 の発表が行われた。日本からは NII, 大阪大学、産業技術総合研究所, DBCLS など研究機関のほか、富士通、日立、東芝、ソニー、Yahoo! などから 15 名ほどの参加があった。参加者の主な専門は、AI, データベース、ソーシャルネットワーク、分散コンピューティング、Web 技術、情報システム、HCI, 自然言語処理、社会科学など多岐にわたり、会議の主なトピックも Linked Data の品質、検索、生命科学への応用、データ統合、検索、質問応答、クエリ言語 SPARQL, オントロジーに基づくデータアクセス、クエリ書換え、推論、自然言語処理、情報抽出、ユーザインタラクション、パーソナライゼーション、ソーシャルメディア、オントロジーアライメント、センサ情報、ストリーム情報処理などさまざまである。そのため、いくつかの異なる背景をもった研究者同士がお互いの知恵と技術をもち合い、それまでに見つかっていなかった新しい問題を相互に研鑽しあいながら解決していく様子が見られた点が、本会議での大変印象的な部分の一つであった。

会議の構成は、Research Track, In-Use Track, Poster & Demo Track, Industry Track などからなる (ただし、

*1 iswc2014.semanticWeb.org

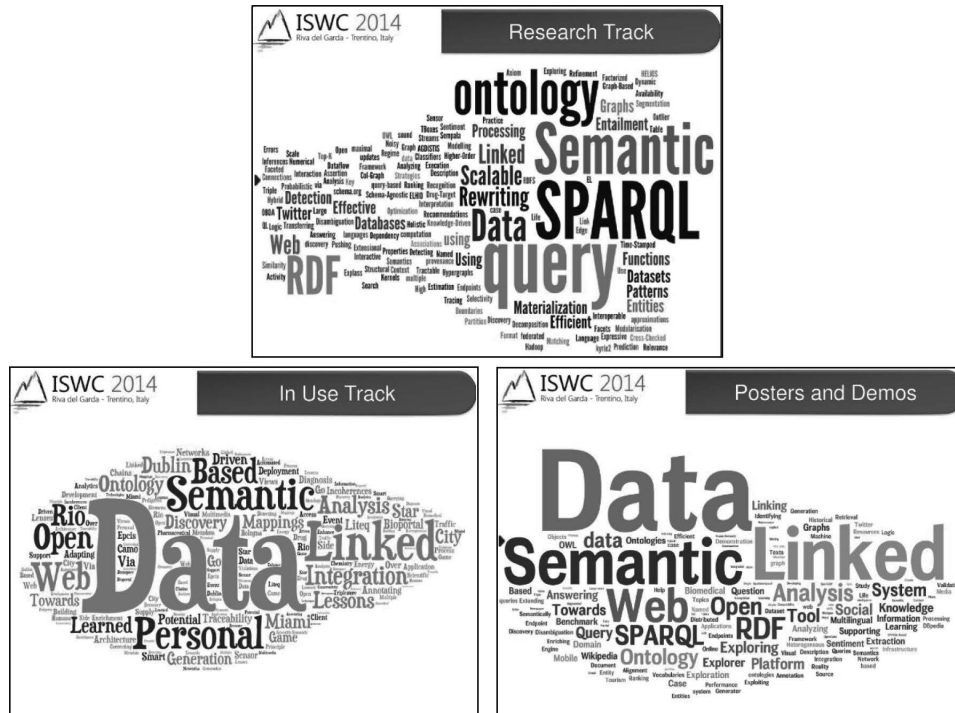


図1 Track 別タグクラウド

各セッションはこれらの Track を横断して、トピックごとにまとめられている)。Research Track には 180 編の投稿があり、38 編（採択率 21%）が採択された。また、In-Use Track は 46 編中、15 編（33%）が採択された。Poster & Demo Track は、前年比 50% 増の 156 編もの投稿があり、71 編がポスターに、50 編がデモとして選ばれた。Industry Track は British Telecom, IBM, Oracle, Siemens をはじめ、ベンチャー企業などから 39 編の投稿があり、7 編（18%）がフルペーパーとして、23 編がショートペーパーとして採択された。さらに、八つのチュートリアル、23 のワークショップが併催された。特に、2014 年はソフトウェア開発者向けデベロッパーワークショップが新設され、Linked Data やセマンティック技術における実装上の問題に対する解決策、方法論などについて議論が行われた。また、ISWC では恒例となっている Semantic Web Challenge も今年で 12 回目を迎え、セマンティック Web の実現に向けた現実的な進展について 15 件のデモが行われた(2・12 節参照)。また、ベストペーパーとしては、それぞれ以下の論文が選ばれた。なお、ベストポスターとベストデモは参加者の投票によって選ばれたものである。

- 10-years award
The Protege OWL Plugin: An Open Development Environment for Semantic Web Applications [Knublauch 04]
- Best paper research track
AGDISTIS-Graph-Based Disambiguation of Named Entities using Linked Data [Usbeck 14]

- Best in use paper
Semantic Traffic Diagnosis with STAR-CITY: Architecture and Lessons Learned from Deployment in Dublin, Bologna, Miami and Rio [Lecue 14]
 - Best poster
Extracting Architectural Patterns from Web data [Gadiraju 14]
 - Best demo
Low-Cost Queryable Linked Data through Triple Pattern Fragments[Verborgh 14b]
 - Semantic Web Challenge-Big Data Track
Extending Tables with Data from over a Million Websites*2
 - Semantic Web Challenge -Open Track
Mining the Web of Linked Data with RapidMiner *2
- 投稿論文のタグクラウド（図1）を見ると、Research Track は SPARQL, Semantic, Query, Ontology, Data, RDF などが同程度の大きさであるが、他のトラックはいずれも Data が圧倒的に大きくなっている。報告者の所感だが、全体的な傾向として昨今のビックデータへの注目からか、従来のオントロジーや推論技術から軽量のセマンティクスを用いたデータ分析にフォーカスが移っているように思われる。ビックデータには三つの V (Volume, Velocity, Variety) があるといわれるが、ビックデータ分析のうち、60%の時間を費やすのは Variety, あるいは Heterogeneity などに起因する data

*2 challenge.semanticWeb.org/2014/submissions

preparation であるといわれている [NASA 12]. こうした問題こそ、これまで本分野で中心的に扱われてきたトピックである. 一方で、良い研究として認められるためには大量のデータを対象とすることが前提条件となってきた. そのため、一般の研究者にとっては分析対象となるデータをいかに入手するかも重要である. また、欧州ではセマンティック技術に多くのファンドを付けた FP7 が終了し、新たに始まった Horizon 2020 ではビッグデータ、ビジネス、イノベーションが期待されていると聞く. 今後、ビッグデータ分析へのセマンティクスの活用が一層盛んになるだろう. なお、ISWC 2015 は 10 月 11 ~ 15 日に米国ペンシルベニア州 Bethlehem にて開催予定である.

以下の節では、著者らが最近の研究動向として興味深く感じた講演、セッション、ワークショップについて簡単に紹介する.

2.2 Keynote Talk: Web Search — From the Noun to the Web

Google の Vice President である Raghavan 氏によるこの招待講演では、まだ Google 検索が生まれる前の検索の世界の話から、今日の、ユーザが能動的に検索をしなくてもユーザのいる場所や時間帯やスケジュールなどに応じて自動的に検索が行われ、その結果がユーザの手元の端末にさりげなく表示される世界（講演者はこれを Proactive で Context を生かした検索と呼んでいた）に至るまでの過程を、非常に簡潔にまとめた。Knowledge Graph のような、より最新の先端的研究の成果についての話題は残念ながらこの講演内では出なかったが、オープンで意味・構造をもったデータが実際の検索エンジンの場面でも効果的に使われている様子をその基礎のアルゴリズムから丁寧にまとめた本講演は、聴講した学生参加者達にとって有益なものであったであろうと感じられた。著者らには、一見すると高度で計算的な負荷も高そうに見える Context を踏まえた Proactive な情報の提示（例えば、GoogleNow^{*3} など）は、検索エンジンへの負荷を低減・制御可能にするという意味でむしろ有益であるという講演者の指摘が大変興味深かった。

2.3 Keynote Talk: Semantic Challenges in Getting Work Done

南カリフォルニア大学情報科学研究所の Associate Director で、同大学コンピュータサイエンス学部研究教授の Yolanda Gil 先生からは Semantic Challenges in Getting Work Done と題したキーノートが行われた。講演では、科学研究（特にデータ分析）の生産性を向上させるための Semantic Challenge として、Todo 管理（個

人の Todo 管理および複数人の Todo 調整など）、豊富な知識が必要となるタスク（セマンティックワークフローや自動論文生成など）、共同タスク（Email-less なプロジェクトの調整やオープンサイエンスプロセスのためのフレームワークなど）を支援するためのさまざまな研究が紹介された。

2.4 Keynote Talk: The Semantic Web in an Age of Open Data

Industry Track からは、ODI (Open Data Institute, イギリスが 2011 年 11 月に設立したオープンデータを活用したビジネスを支援する研究所)^{*4} の共同創設者兼議長である Southampton 大学教授 Nigel Shadbolt 先生によるキーノートが行われた。講演では、過去 5 年間で政府の支出や犯罪、裁判記録、教育、健康、交通、地理情報、環境など多岐にわたるデータが公開されてきたことをあげ、これからは、皆が Author, Scientist になれる Open Innovation の時代であるとされた。また、今後はセマンティクスの活用がオープンデータの価値と利用率を上げるとし、成功事例や普及の障害などが紹介された。

2.5 Session: Querying

本セッションでは、大量の Linked Data やそれを扱うオントロジーに対する効率的な問合せ手法に関するアイデアが発表された。

データベース分野などで一般的な問合せ処理では、データそのものに対していかに効率的な処理が実現可能かを議論することが多いが、Linked Data ではデータの不完全性を考慮して、その処理結果の完全性に対する要求をある程度柔軟に設定することで、例えば問合せ対象のデータそのものの一部のみへ問合せ処理の対象を限定することで、その処理を現実的な処理負荷で実行可能なものにするアプローチも考えられる。そのアプローチの一つとして、Rietveld ら [Rietveld 14] は、大規模な RDF トリプルをうまくサンプリングする方法について提案した。アイデアとしては、ページランクに類似した比較的単純なテクニックを使ったものであるが、総量 1.4 trillion トリプル、計 1500 データセットに対する横断的な評価を緻密に行っている点で興味深く、その規模でもスケールするように Hadoop で分散処理するなどの工夫がなされた点が一つの特長となっていた。

Wagner ら [Wagner 14] は、Hybrid-query と呼ぶアンテーションなどで生じるメタ情報中の比較的長いテキストデータから、キーワード検索を伴うクエリを効率的に実行する手法が提案された。具体的な利用場面としては、例えば DBpedia 内のデータに対するクエリの一部で、テキストデータとして書かれた情報に対するキーワードマッチングを含むような操作が行われる場合が想定され

*3 www.google.com/landing/now

*4 opendatainstitute.org

る。どこにどのキーワードやトピックに関連するデータが入っているかを事前にペイジアンネットワークをつくっておいて予測可能にすることで、ある程度大規模な処理に対してもスケールするというメリットを実験結果から述べていた。一方で、Hybrid-query 自体へのニーズの多寡については、セッション参加者の間でも議論となった。

Butt ら [Butt 14] は、オントロジー自身を検索するという課題に対してのアプローチを示しており、もともとランキングなどが難しいとされる「ある概念を含む適切なオントロジー」を探す問題に対して、BM25 やページランクなどの既存の文書検索などに有効な手法がどの程度有効に機能するかを、大規模オントロジーデータセットを対象に評価している。結論として古典的な TF/IDF が最も安定して良い結果が出ていたという点には、プラクティカルなシステムの実装を考える人達にとっても見るべき結果であるだろう。

2.6 Session: Reasoning

本セッションでは、セマンティック Web における重要なトピックの一つである推論 (Reasoning) に関する理論的背景となっている記述論理 (Description Logics) に関する処理を効率的に処理・解析する方法についての発表が行われた。

Carral ら [Carral 14] は、推論の体系の一つである SROIQ やそれより狭い範囲である Horn-SROIQ よりもさらに範囲を狭めた RSA という推論の範囲を規定して、これを DATALOG ベースの Reasoner で高速に推論するというアイデアを提案していた。既存の高速な Reasoner として知られる Hermit^{*5} よりも高速にかつ省メモリで推論できることが大きなメリットである反面、連言的質問応答 (conjunctive query answering) は今後の課題であるとした。

Console ら [Console 14] は、オントロジーを適度に「端折って (approximate して)」しまうことで、OWL2 オントロジーを OWL2-QL に相当するオントロジーに落とし込むというアイデアが示された。すでに、AAAI 2007 に OWL-DL を “approximate” するアイデアは提案されているが、本発表では k -approximate、すなわち最低限 k の公理 (axiom) について正しく推論できるという考え方を採用している点がポイントであると述べられていた。単に近似という意味での “approximation” というとその推論の正しさが保証されなくなる点が心配されるが、提案された “approximation” では発表者らのいう “minimal change” の原則により健全性 (soundness) 自体はできるだけ残す方向の処理が行われていると述べられていた。

Glimm ら [Glimm 14] は、特にオントロジーに基づく

データアクセス (OBDA) のドメインでの利用を想定して、同じような推論を複数回行うことを避けるために、A-Box を抽象化した abstract ABox というものを追加するアイデアが提案された。この方法は、QA システムでもよくあるマテリアライズ (キャッシュ) の考え方に近いが、A-Box における推論に特化したコンパクトな抽象化が実現できている点に特長が見られた。

Kazakov ら [Kazakov 14] は、OWL-EL の範囲で推論の過程をトレース可能にする試みが述べられた。Prolog のような単純な推論過程をもつ場合には、そのバックトレースを行うことはそれほど難しくないように考えられるが、OWL-EL の推論では推論のフローが直線的になるとは限らず現状ではそのような仕組みも必ずしもないため、その現実的な実現方法を検討した点が興味深く感じられた。ただ、現状ではこの研究の適用範囲は OWL-EL の範囲にとどまっており、現状のアプローチは disjunction を含めた推論へ発展させることは難しい、と発表者は質疑応答の際に述べていた。

2.7 Session: SPARQL Extension

本セッションでは、Linked Data に対する問合せの標準である SPARQL の拡張の可能性を議論する発表が行われた。

Buil-Aranda ら [Buil-Aranda 14] は、Federated query の実行戦略に関する提案で、Federated query がタイムアウトしてしまったような場合に現状の SPARQL の使用では NULL rejection がなされないことから、結果に問題が生じてしまう点に対して、個々のサブクエリの実行の制約方法を FILTER や UNION などを用いてクエリを書き換えることでどの程度抑制できるかを、実際の著名な SPARQL 処理系である Jena-Fuseki^{*6}、Sesame^{*7}、Virtuoso^{*8} の三つについて詳細に検討して調査した結果が報告された。単に現状の SPARQL の仕様上の問題を指摘する意図なのかという問いには、現状の見つかっている課題に対してどのような対処方法があるかを実際に調べてみたというものであり、現状の仕様を批判する意図はないと回答されていた。

Atzori [Atzori 14] は、SPARQL にある組込み Function の少なさや非互換性への対処方法として、「Web of functions」として Web 上から accessible に URI で参照可能な関数を定義・実装し、それらをユニバーサルに SPARQL から利用できるようにしようという試みについて述べられた。実際の実装にあたっては、単純な call を外部に毎回行っていると非常にオーバーヘッドが大きいため、動作速度が重要な場合は proxy endpoint を使うことで対処するなどのアイデアも併せて示された。

*6 jena.apache.org/documentation/serving_data

*7 rdf4j.org

*8 virtuoso.openlinksw.com

*5 hermit-reasoner.com

当然、クエリ処理中の外部関数呼出しに関しては、それに対するデータの外部漏えいなどが心配になる部分もあるが、すでに WebAPI などを使うことに慣れている現状もあり、オープンなデータに対する処理であれば、ある程度寛容に考えてよいのではないかというのが発表者の意見であった。アイデアとしては非常に単純に思えるが、その実現を実際の SPARQL などの仕様・文法の枠内でうまく行おうとすると、いくつかの工夫がなされているように感じられた。

2.8 Session: Large-scale RDF Processing and Dataset Availability

本セッションでは、主に大量の Linked Data を扱う場合における処理について、問合せ処理の効率化以外の側面を考慮したいいくつかのアプローチが提案された。

Maali ら [Maali 14] は、SPARQL では直接の扱いがされない「二つのグラフ構造同士の Join」などを大量のデータに対して行える処理系としての SYRql と、その理論的基盤としての RDF Algebra を提案している。実装では JSON-LD^{*9} を使っており、140 million トリプルのデータを使った検証による評価が示されている。一見すると通常の SPARQL にもある通常の Join 演算でも同じことができそうに見え、会場でも同種の質問がされたが、本研究では RDF から一部を切り出す前のグラフ構造を崩さずにこの演算を高速に実行できる点が特長であると主張していた。

Schätzle ら [Schätzle 14] は、SQL-on-Hadoop といった流れに乗って、SPARQL でも同種の実行をできるようにしながら、それをインタラクティブに簡単に制御・実行できるようにしたシステムを提案していた。実装としては、SPARQL から ImparaSQL^{*10} への変換を行うようになっていて、10 台程度のクラスタをつくった実験にとどまる点と、サポートできる範囲がまだ SPARQL1.0 にとどまる点が今後の課題となっている。Vertica^{*11} のようなカラム型 DB を併用した場合やそれとの比較については、会場からもその質問があったが、現状ではまだその比較検討を行えていないようであった。こうしたデータベース分野の技術の転用については、アイデア自体にはそれほど新しさを感じないものもあるが、その実際の性能に対する丁寧な検証が行われていたり、そうして構築した基盤がオープンになっていたりする点が、この会議の一つの特徴的な点であるように報告者には感じられた。

Verborgh ら [Verborgh 14a] は、SPARQL のエンドポイントやそれに対応したクライアントを、RDF トリプ

ルの断片 (fragment) を集めてつくれるようにするフレームワークを提案した。SPARQL 1.1 以降では横断的クエリもすでに実現されるが、DBpedia などの膨大なデータ量の LOD データの中から必要な部分だけを上手に切り出して一つの SPARQL エンドポイントを構成するための方法論の必要性には同意できる。本研究では、具体的なフォーム型のクライアントを起点として、そのクエリから具体的にどのような LOD に対してのアクセスがなされるかを計測し、そこで頻繁なアクセスのある部分を切り出すというアイデアに基づいて、この方法を実現している。複雑なクエリに対する処理などの対処にはまだ課題があるとのことであったが、データのオープン性を生かし、必要なデータをアプリケーションのために切り出してくるという発想自体は興味深いと感じられた。

Aluç ら [Aluç 14] は、RDF のクエリエンジンに対する負荷テストの方法論に関する考察を述べていた。一般のベンチマーク問題では平均的な性能を調べられるものの、実際の RDF クエリ実行処理系では特定の「苦手な」クエリがあり、できるだけそうしたクエリに上手に触れるようなテストパターンを用意することで、クエリ実行時に極端な実行時間の増加に伴ってクエリの実行エンジンが落ちてしまうような状況をテストできるとしている。こうした実行特性は、もちろん RDF に限らず一般の RDBMS にも見られるものであるが、LOD として公開されるデータに対する RDF クエリエンジンでは任意のクエリを外部から受け取る可能性があるため、DB アプリケーションにありがちなクエリの書き方で対処するという方法論が必ずしも適用できず、こうした研究の必要性も出てくることになる。

2.9 Session: Linked Data and Data Quality

本セッションでは、Linked Data の検索および視覚的な分析を可能にするツールや、Linked Data の質を評価したり、誤りを訂正するための手法などが提案された。

Sabol ら [Sabol 14] は、SPARQL などのセマンティック技術の背景知識がないユーザでも Linked Data の検索および視覚的な分析を可能にするツール (Query Wizard^{*12} および Vis Wizard^{*13}) を提案している。Query Wizard は、標準的な Web 検索エンジンと同程度の簡単な操作で Linked Data の検索を行うことが可能なツールで、表形式のインタフェースを提供している。Vis Wizard は、Linked Data の視覚化およびインタラクティブな分析を可能にするツールである。

自動構築されたオントロジーや Linked Data には、一定の誤りが含まれる可能性があり、それらの質の評価や誤りの発見は課題となっている。[Florion 14] では、ク

*9 json-ld.org

*10 www.cloudera.com/content/cloudera/en/documentation/cloudera-impala/latest/topics/impala_langref.html

*11 www.vertica.com

*12 code.know-center.tugraz.at/search

*13 code.know-center.tugraz.at/vis

クラウドソーシングを用いてオントロジーにおけるクラスがある対象領域としてふさわしいかどうかを判定したり、is-a 関係が正しいかどうかを判定する Protege プラグインを提案している (ISWC 2014 のポスターとデモセッションでも同様の発表があった)。[Zaveri 13] では、クラウドソーシングを用いて汎用的に Linked Data の質を評価するツール TripleCheckMate^{*14} を提案し、DBpedia におけるトリプルの評価を行っている。

同様の背景により、本セッションでは Linked Data における誤りを自動的に発見するための研究発表があった。Fleischhacker ら [Fleischhacker 14] は、外れ値の発見手法を用いて Linked Data における数値 (人口や身長プロバティ値) の誤りを発見する手法を提案している。Zhu ら [Zhu 14] は、Adaboost with C 4.5 を用いたタイプアサーションペアのマルチクラス分類により、トリプルにおけるリソースのタイプの誤り (例えば、人間と場所クラスの両方をタイプとしてもつリソースなど) を発見する手法を提案している。

2.10 Session: NLP & IEs

本セッションでは、Web ページなどの非構造化データからの構造化 (RDF 化, Linked Data 化) に対するアプローチなどが提案された。

2.1 節「全体概要」で紹介したベストペーパー AGDISTIS [Usbeck 14] は、固有名詞 (対象は、人、組織、場所) の曖昧性解消 (Disambiguation) を扱ったものであり、ラベル拡張と文字列類似度で候補となる語のグラフを作成し、スケーラビリティ向上のため HITS (Hypertext-Induced Topic Search) を用いて探索している点に特徴がある。本分野では、ベクトル空間とコサイン類似度に基づく DBpedia クラスへの対応付けを行う DBpedia Spotlight [Mendes 11] がベンチマークとしてよく使われるが、それらがベースとなる KB を既知としているのに対し、それを事前に知らないもの (agnostic) としている点も特徴である。多言語対応、データ・セットの大きさ、Web サービス公開なども大きなポイントとなっている。

2.11 Session: Sensors

本セッションでは、昨今、注目を集めている M2M (Machine to Machine) や IoT (Internet of Things) におけるセンサ情報にセマンティクスを付与するアプローチが提案された。

2.1 節「全体概要」でも紹介したベストペーパー STAR-CITY [Lecue 14] は、交通状況の分析と推定を目的とした IBM のシステムである。アイルランドのダブリンで行った Dublinked^{*15} をイタリアのポローニヤ、アメリ

カのマイアミ、ブラジルのリオへ展開し、より汎用的にしたアーキテクチャの提案とアプリケーション事例である。過去および現在の交通状況の提供や、バス情報やソーシャルメディア、工事情報、イベント情報 (音楽、政治的)、乗降者数などのデータを統合することによる効率的な移動計画の作成をサポートしている。また、コンテキスト取得のためにセンサ情報にセマンティクスを付与している。2012 年の Semantic Web Challenge で優勝した研究の発展版だが、今回は主にバスにフォーカスした点に特徴がある。

また、オーストラリアの CSIRO (Commonwealth Scientific and Industrial Research Organisation) からは、現状、各農家、農業法人単位で活用されているセンサデータを国として統合して活用するために、メタデータを付与して病害虫などへの対策を進める Agricultural decision support systems の構築について発表があった。ここでは、タスマニアのぶどう栽培において、W3C SSN Ontology [Compton 12]などをベースにセンサデータに対してメタデータを付与し、それを検索することで、場所、直近のキャリブレーション、一定期間のデータ品質といった条件に沿った適切な情報の選択、順位付けを行っている。

2.12 Session: Semantic Web Challenge

Semantic Web Challenge は、セマンティック Web 技術を用いてエンドユーザ向けアプリケーションを構築することにより、セマンティック Web 技術の可能性を示すことを目的としている。特に制限のない Open Track と大規模データセットの利用が必須となる Big Data Track の二つのトラックから構成され、両トラックともに実際にエンドユーザ向けのアプリケーションを構築し、デモを行うことが必須条件となっている。今年は 15 件の投稿があり、ポスターとデモセッションにおいて審査員に対してデモを行い、7 件 (Open Track 6 件と Big Data Track 1 件) がファイナリストとして口頭発表に選ばれた。

日本からは山田らがファイナリストに選ばれ、Web ページ中のエンティティ名に自動的にリンクを張り、ユーザがリンクを選択した際に、エンティティのクラス情報を用いてコンテキストを考慮したエンティティの要約情報をウィジェットに表示する Web ブラウザプラグイン Linkify^{*16} を発表した。

Open Track の最優秀賞に選ばれた Ristoski らの発表では、RapidMiner^{*17} と呼ばれるデータマイニングプラットフォームのための Linked Open Data 拡張^{*18} が提案

*14 github.com/AKSW/TripleCheckMate

*15 www.dublinked.ie

*16 swc14.linkify.mobi

*17 rapidminer.com

*18 dws.informatik.uni-mannheim.de/en/research/rapidminer-lod-extension

された。RapidMinerは、クラスタリング、相関ルール、分類などのデータ分析に利用可能なアルゴリズム、データ入力、フィルタ、視覚化などの機能をGUIを用いて組み合わせることで、データの配置から分析結果の評価までを簡単に行うことが可能なツールである。Linked Open Data 拡張は、RapidMinerにSPARQLエンドポイントまたはローカルRDFファイルからのインポート機能、他のデータセットへ自動的にリンクを張る機能、スキーママッチング機能などを追加することにより、データマイニングツールを用いたLinked Open Dataの分析を容易に行うことを可能にしている。

Big Data Trackの最優秀賞に選ばれたLehmergらの発表では、Mannheim Search Joins Engine (MSJ Engine)^{*19}が提案された。MSJ Engineは、Billion Triples Challenge 2014 Dataset^{*20}、WebDataCommons Micro-data Dataset^{*21}、WebDataCommons HTML Tables Dataset^{*22}、WikiTables Dataset^{*23}の四つのデータセットから約3600万のテーブル(約30億トリプルのインデックス)を抽出し、それらを用いてユーザが入力したローカルテーブルの拡張を行うことが可能なツールである。例えば、書籍名のみを含むローカルテーブルを入力することで、各書籍に関連する著者、評価、ISBN、出版社などの列を他のデータセットから抽出したテーブルから自動的に抽出し、ローカルテーブルを拡張することが可能となる。

2.13 Natural Language Interfaces for Web of Data Workshop

今回が第1回だが、参加者は42名と活気のあるワークショップであった。主要テーマはLinked Dataに対する質問応答の実現であり、本分野におけるホットピックの一つである。多くの手法は、基本的には言語解析した結果をLODの部分グラフとマッチさせるものであるが、旧DERI (Digital Enterprise Research Institute, 現在は2013年に設立されたInsight Centre for Data Analyticsの一部となっている)からの招待講演Talking to Your Data Natural Language Interfaces for the Schemaless World^{*24}では、Treo QA System^{*25}で用いられているschema-agnostic (事前にデータのスキーマがわからない場合を指す、オープンデータの検索でしばしば想定される)の際に、多次元ベクトルとのコサイン類似度を多段にとって意味的な類似度を求める手

法DSM (Distributional Semantic Model)^{*26}が提案されており、興味深かった。

2.14 Semantic Sensor Network Workshop

IoTやSmart Cityに向けたセマンティクスの活用がメインのワークショップであり、大変活況であった。招待講演Dynamic Semantics for Semantics for Dynamic IoT Environments^{*27}では、IoTへの流れとして、RFID → Wireless Sensor → Smart Device → M2Mを示したうえで、センサデータはWeb上のデータと異なり、(1)センサやセンサネットからのイベントベースであること、(2)時間や空間との関連があり、動的な性質をもっていること、(3)電力やメモリなどネットワークなどによる物理的な制約を受けることをあげた。また、FP7のプロジェクトCity-Pulseで検討されたSmart Cityにおけるユースケースシナリオ^{*28}なども紹介され、セマンティクスの活用はデータ統合、相互運用性の鍵となるが、良いモデルがあるだけではなく、どこにどのようにセマンティクスを付けるかなど、エンドユーザやデータ作成者の利用環境をよく考えたツールの整備が重要であるとの見方が示された。会場からは、セマンティクスをセンサ活用のカタパルトとして見る見方と、あくまで裏方の一部にすぎない、との見方がぶつかり、CSVとの本質的な違いについて質問が飛ぶなど、突っ込んだ議論が交わされ興味深かった。

2.15 NLP & DBpedia 2014 Workshop

本ワークショップは、DBpediaと自然言語処理の関連を探究することを目的としている。自然言語処理技術を用いてWikipediaにおける非構造の記事から情報抽出を行うことによりDBpediaの拡張を試みる研究や、DBpediaやLinked Open Dataを用いてテキストやHTML文書から知識の抽出を試みる研究発表などがなされている。一般発表件数についてはフルペーパー1件、ショートペーパー3件と少なかったが、招待講演(Digital) goodies from the ERC Wishing Well: BabelNet, Babelify, video games with a purpose and the Wikipedia bitaxonomyは、立ち見で部屋から聴講者があふれるほど活況であった。招待講演では、主にWikipedia Bitaxonomy^{*29} [Flati 14], Babel-Net^{*30} [Navigli 12], Babelify^{*31} [Moro 14]についての研究紹介がなされた。Wikipedia Bitaxonomyは、Wikipedia記事における定

*19 searchjoins.Webdatacommons.org

*20 km.aifb.kit.edu/projects/btc-2014

*21 Webdatacommons.org/structureddata

*22 commoncrawl.org

*23 Webdatacommons.org/Webtables

*24 www.slideshare.net/andrenfreitas/talking-to-your-datanatural-language-interfaces-for-a-schemaless-world

*25 treo.deri.ie

*26 easy-esa.org

*27 www.slideshare.net/PayamBarnaghi/dynamic-semantics-for-semantics-for-dynamic-iot-environments

*28 www.ict-citypulse.eu/scenarios/ranking

*29 wibitaxonomy.org

*30 babelnet.org

*31 www.babelify.org

義文から抽出した is-a 関係と Wikipedia におけるカテゴリ階層をアライメントすることにより、構築されたタクソノミーである。BabelNet は、Word-Net, Wikipedia, Wiktionary^{*32}, OmegaWiki^{*33}, Wiki-data^{*34}, 多言語 WordNet を統合することにより構築された、大規模多言語百科事典と意味ネットワークである。Babelfy は、BabelNet を利用することにより、多言語テキストに対して、多義性解消と Entity Linking を可能にする。また、テレビゲームを用いることで、意味ネットワークの検証と拡張を行う研究紹介もあった^{*35} [Vannella 14]。ポスターとデモセッションでも、Babelfy についての発表があり、Java 言語により RESTful API を実行する方法の説明などがなされていた。Wikipedia Bitaxonomy, BabelNet, Babelfy は、セマンティック Web のアプリケーションに活用可能な新たな情報資源や Web サービスとして興味深かった。

2.16 Ordering and Reasoning Workshop

本ワークショップでは、セマンティック Web におけるクエリ処理の一つのトレンドである、Top- k クエリ処理を扱った発表が多数を占めた。Top- k クエリ処理では「最も上位にランクされる k 個のもの」のみを効率的に見つける必要があるが、セマンティック Web ではデータの順序付けそのものが一つの課題であり、Gillani ら [Gillani 14] の発表では、OWL-DL では直接扱えない Temporal Data の効率的な扱いを試みるアイデアが提案された。また、クエリ処理の効率化手法そのものではなく、RDF データに対するクエリ処理エンジンに対する効果的なベンチマーク方法に関する提案も、このワークショップのオーガナイザの一人である Zahmatkesh ら [Zahmatkesh 14] から発表された。後日、この Zahmatkesh 氏に本会議のポスターセッションでその詳細を聞いたが、実際には報告者が考えている以上にまだ性能向上の余地が残されているようであり、データベース分野の知識をもった研究者にとっても、挑戦のしがいのあるテーマがここにはまだいくつも残されているように感じられた。

2.17 その他のワークショップの動向

ワークショップでは、特定の技術的な側面にフォーカスして集中的な議論が行われるもの以外にも、実世界にある情報をどのように効率的かつ効果的に扱えるようにするかという課題に対して、具体的なデータやモデルを扱う側とそれらの処理方法を扱う側とが、アイデアをもち寄りながら議論を進めているものもあった。

例えば、オープンデータと行政などとの関わりなどについて扱われたワークショップである SemStats2014 (2nd International Workshop on Semantic Statistics)^{*36} では、イタリアにおける行政の統計 (国勢調査) データをモデルケースとして、歴史の変遷をたどっていきけるようにデータやモデルの由来そのもののモデル化も行っていく必要性 [Lodi 14] についての議論がなされると同時に、別の発表ではそうした大量な統計データを効率的に処理・検証するためのシステムの具体的な実装例 (例えば、Emilio ら [Emilio 14] など) が提示されながら、そうした技術の実世界での利用の難しさの解消についての熱心な議論が行われていた。

こうした多様なワークショップでの参加者による熱のこもった議論の成果は、そうした議論の場をつくり育てていくことにもつながっており、そうした議論の場に参加することで、自らがそのコミュニティづくりに参加していくことの動機付けが自然と行われていく印象があった。著者らのコミュニティからもこうしたコミュニティづくりの場面へのさらなる積極的な参加が期待される。

3. JIST 2014 参加報告

JIST 2014 (4th Joint International Semantic Technology)^{*37} は 2014 年 11 月 9 ~ 11 日の間、タイの Chiang Mai で開催された。JIST は、アジア・太平洋地区を中心としたセマンティック技術に関する国際会議であり、以前に ASWC (Asian Semantic Web Conference) と呼称していたものが名前を変えた会議である。4 回目 (ASWC 時代から数えると 8 回目) となる今回は 15 か国から 100 名近い参加者があり、フルペーパー、ショートペーパー、ポスターを合わせて約 40 件の発表が行われた。

JIST は ISWC のアジア版といった位置付けであるが、データ分析への極端な偏りはなく、AI や Inference, Social Web といったキーワードが広く散見される。また、日本の研究コミュニティの貢献も大きく、General Chair を務められた山口高平先生 (慶應義塾大学) はじめ、Organize Committee に 3 名、Program Committee に 13 名の日本人研究者が名を連ねている。また、日本からの投稿も大変活発な会議である。なお、JIST 2015 は 2015 年 11 月 11 ~ 13 日に中国湖北省宜昌市にて開催予定である。

4. おわりに

本稿では、ISWC 2014 参加報告を中心とし、海外におけるセマンティック技術の研究動向について著者らの主観に基づいて報告した。本分野は、これまで国内では

*32 www.wiktionary.org

*33 www.omegawiki.org

*34 www.wikidata.org

*35 knowledgeforge.org

*36 <http://semstats2014.wordpress.com>

*37 language-semantic.org/jist2014

人工知能学会セマンティックウェブとオントロジー研究会^{*38}や、人工知能学会全国大会オーガナイズドセッション「Linked Data とオントロジー」^{*39}、Linked Open Data チャレンジ Japan^{*40}などの関係者が中心に活動してきた。しかし、昨今では国内でもオープンデータの活用がさまざまな分野で検討されてきたことで、これまでセマンティック関連技術に関わって来られなかった方々も多く参入してきている。DBpedia / DBpedia Japanese を活用して、データ分析やマイニングについて研究、実践されている方々も同様に数多く参入してきている。ISWC では、大規模データベースや自然言語処理など他分野から参入されて、良い研究成果を上げられているグループがいくつも存在する。

国内においても、ぜひ、他分野からの視点で本分野に新しい知見をもち込んでいただくことに期待したい。特に、ISWC 2016 神戸開催に向けては、従来のコミュニティの枠にとらわれず、広く日本の研究者、技術者からの参加と会議成功に向けたご協力をお願いしたいと考えている。

◇ 参 考 文 献 ◇

- [Aluç 14] Aluç, G., Hartig, L., Ozsu, M. T. and Daudjee, K.: Diversified stress testing of RDF data management systems, *Proc. ISWC 2014*, LNCS 8796, pp. 197-212 (2014)
- [Atzori 14] Atzori, M.: Toward the web of functions: Interoperable higher-order functions in SPARQL, *Proc. ISWC 2014*, LNCS 8797, pp. 406-421 (2014)
- [Buil-Aranda 14] Buil-Aranda, C., Polleres, A. and Umbrich, J.: Strategies for executing federated queries in SPARQL1.1, *Proc. ISWC 2014*, LNCS 8797, pp. 390-405 (2014)
- [Butt 14] Butt, A. S., Haller, A. and Xie, L.: Ontology search: An empirical evaluation, *Proc. ISWC 2014*, LNCS 8797, pp. 130-147 (2014)
- [Carral 14] Carral, D., Feier, C., Grau, B. C., Hitzler, P. and Horrocks, I.: Pushing the boundaries of tractable ontology reasoning, *Proc. ISWC 2014*, LNCS 8797, pp. 148-163 (2014)
- [Compton 12] Compton, M., et al.: The SSN ontology of the w3c semantic sensor network incubator group, *J. of Web Semantics*, Vol. 17, pp. 25-32 (2012)
- [Console 14] Console, M., Mora, J., Rosati, R., Santarelli, V. and Savo, D. F.: Effective computation of maximal sound approximations of description logic ontologies, *Proc. ISWC 2014*, LNCS 8797, pp. 164-179 (2014)
- [Emilio 14] Emilio, J., Gayo, L., Farhan, H., Fernández, J. C. and Rodriguez, J. M. A.: Representing verifiable statistical index computations as linked data, *Proc. 2nd Int. Workshop on Semantic Statistics* (2014)
- [Flati 14] Flati, T., Vannella, D., Pasini, T. and Navigli, R.: Two is bigger (and better) than one: The wikipedia bitaxonomy project, *Proc. 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, pp. 945-955 (2014)
- [Fleischhacker 14] Fleischhacker, D., Paulheim, H., Bryl, V., Völker, J. and Bizer, C.: Detecting errors in numerical linked data using cross-checked outlier detection, *Proc. ISWC 2014*, LNCS 8797, pp. 357-372 (2014)
- [Florian 14] Hanika, F., Wohlgenannt, G. and Sabou, M.: The uComp protege plugin: Crowdsourcing enabled ontology engineering, *Proc. 19th Int. Conf. on Knowledge Engineering and Knowledge Management (EKAW 2014)*, pp. 181-196 (2014)
- [Gadiraju 14] Gadiraju, U., Kawase, R. and Dietze, S.: Extracting architectural patterns from web data, *Proc. ISWC 2014, Posters & Demonstrations Track, CEUR-WS*, Vol. 1272 (2014)
- [Gillani 14] Gillani, S., Picard, G., Laforest, F. and Zimmermann, A.: Towards efficient semantically enriched complex event processing and pattern matching, *Proc. 3rd Int. Workshop on Ordering and Reasoning (OrdRing 2014)*, CEUR-WS, Vol. 1303, pp. 47-54 (2014)
- [Glimm 14] Glimm, B., Kazakov, Y., Liebig, T., Tran, T. K. and Vialard, V.: Abstraction refinement for ontology materialization, *Proc. ISWC 2014*, LNCS 8797, pp. 180-195 (2014)
- [Kazakov 14] Kazakov, Y. and Klinov, P.: Goal-directed tracing of inferences in EL ontologies, *Proc. ISWC 2014*, LNCS 8797, pp. 196-211 (2014)
- [Knublauch 04] Knublauch, H., Fergerson, R. W., Noy, N. F. and Musen, M. A.: The protege OWL plugin: An open development environment for semantic web applications, *Proc. ISWC 2004*, LNCS 3298, pp. 229-243 (2004)
- [Lecue 14] Lecue, F., Tucker, R., Tallewi-Diotallewi, S., Nair, R., Gkoufas, Y., Liguori, G., Borioni, M., Rademaker, A. and Barbosa, L.: Semantic traffic diagnosis with STAR-CITY: Architecture and lessons learned from deployment in Dublin, Bologna, Miami and Rio, *Proc. ISWC 2014*, LNCS 8797, pp. 292-307 (2014)
- [Lee 01] Lee, T. B., Hendler, J. and Lassila, O.: The semantic web, *Scientific American*, pp. 29-37 (2001)
- [Lodi 14] Lodi, G., Maccioni, A., Scannapieco, M., Scanu, M. and Tosco, L.: Publishing official classifications in linked open data, *Proc. 2nd Int. Workshop on Semantic Statistics* (2014)
- [Maali 14] Maali, F., Ravindra, P., Anyanwu, K. and Decker, S.: SYRql: A dataflow language for large scale processing of RDF data, *Proc. ISWC 2014*, LNCS 8796, pp. 147-163 (2014)
- [Mendes 11] Mendes, P. N., Jakob, M., Garcia-Silva, A. and Bizer, C.: DBpedia spotlight: Shedding light on the web of documents, *Proc. 7th Int. Conf. on Semantic Systems (I-Semantics 2011)*, pp. 1-8 (2011)
- [Moro 14] Moro, A., Raganato, A. and Navigli, R.: Entity linking meets word sense disambiguation: A unified approach, *Trans. Association for Computational Linguistics (TACL)*, Vol. 2, pp. 231-244 (2014)
- [NASA 12] NASA: A.40 computational modeling algorithms and cyberinfrastructure (Dec. 19, 2011), Technical Report, National Aeronautics and Space Administration (NASA) (2012)
- [Navigli 12] Navigli, R. and Ponzetto, S. P.: BabelNet: The automatic construction, Evaluation and application of a wide-coverage multilingual semantic network, *Artificial Intelligence*, Vol. 193, pp. 217-250 (2012)
- [Rietveld 14] Rietveld, L., Hoekstra, R., Schlobach, S. and Guéret, C.: Structural properties as proxy for semantic relevance in RDF graph sampling, *Proc. ISWC 2014*, LNCS 8797, pp. 81-96 (2014)
- [Sabol 14] Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B. and Granitzer, M.: Discovery and visual analysis of linked data for humans, *Proc. ISWC 2014*, LNCS 8797, pp. 309-324 (2014)
- [Schätzle 14] Schätzle, A., Przyjaciół-Zablocki, M., Neu, A. and Lausen, G.: Sempala: Interactive SPARQL query processing on hadoop, *Proc. ISWC 2014*, LNCS 8796, pp. 164-179 (2014)
- [Usbeck 14] Usbeck, R., Ngomo, A. C. N., Roder, M., Gerber, D., Coelho, S. A., Auer, S. and Both, A.: AGDISTIS-graph-based disambiguation of named entities using linked data, *Proc. ISWC 2014*, LNCS 8796, pp. 457-471 (2014)
- [Vannella 14] Vannella, D., Jurgens, D., Scarfini, D., Toscani, D. and Navigli, R.: Validating and extending semantic knowledge bases using video games with a purpose, *Proc. 52nd Annual*

*38 sigsw.org

*39 www.ai-gakkai.or.jp/jsai2015/os#os-8

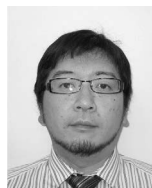
*40 lod.sfc.keio.ac.jp/challenge2014

Meeting of the Association for Computational Linguistics (ACL 2014), pp. 1294-1304 (2014)

- [Verborgh 14a] Verborgh, R., Hartig, O., Meester, B. D., Haesendonck, Vocht, G. L. D., Sande, M. V., Cyganiak, R., Colpaert, P., Mannens, E. and Van de Walle, R.: Querying datasets on the web with high availability, *Proc. ISWC 2014*, LNCS 8796, pp. 180-196 (2014)
- [Verborgh 14b] Verborgh, R., Hartig, O., Meester, B. D., Haesendonck, G., Vocht, L. D., Sande, M. V., Cyganiak, R., Colpaert, P., Mannens, E. and Walle, R. V. D.: Low-cost queryable linked data through triple pattern fragments, *Proc. ISWC 2014*, Posters & Demonstrations Track, CEUR-WS, Vol. 1272 (2014)
- [Wagner 14] Wagner, A., Bicer, V., Tran, T. and Studer, R.: Holistic and compact selectivity estimation for hybrid queries over RDF graphs, *Proc. ISWC 2014*, LNCS 8797, pp. 97-113 (2014)
- [Zahmatkesh 14] Zahmatkesh, S., Valle, E. D., Dell'Aglia, D. and Bozzon, A.: Towards a Top-K SPARQL query benchmark generator, *Proc. 3rd Int. Workshop on Ordering and Reasoning (Or-dRing2014)*, CEUR-WS, Vol. 1303, pp. 47-54 (2014)
- [Zaveri 13] Zaveri, A., Kontokostas, D., Sherif, M. A., Bühmann, L., Morsey, M., Auer, S. and Lehmann, J.: User-driven quality evaluation of DB-pedia, *Proc. 9th Int. Conf. on Semantic Systems*, pp. 97-104 (2013)
- [Zhu 14] Zhu, M., Gao, Z. and Quan, Z.: Noisy type assertion detection in semantic datasets, *Proc. ISWC 2014*, LNCS 8797, pp. 373-388 (2014)

2015年5月27日 受理

著者紹介



川村 隆浩 (正会員)

1994年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年、株式会社東芝研究開発センター入社。2001～02年米国カーネギー・メロン大学ロボティクス研究所客員研究員兼任。2003年より電気通信大学大学院情報システム学研究科客員准教授兼任。2007年より大阪大学大学院工学研究科非常勤講師兼任。2015年より科学技術振興機構情報企画部情報分析室主任調査員。現在に至る。博士(工学)。2012年 ISWC 10-Year Award。2014年本学会研究会優秀賞。主としてセマンティック Web、エージェント技術の研究・開発に従事。情報処理学会会員。



森田 武史 (正会員)

2003年静岡大学情報学部情報科学科卒業。2005年同大学院情報学研究科修士課程修了。2007年日本学術振興会特別研究員(DC2)。2008年慶應義塾大学大学院理工学研究科後期博士課程修了。同年、日本学術振興会特別研究員(PD)。2009年慶應義塾大学大学院理工学研究科特別研究助教。2011年青山学院大学社会情報学部助手。2014年同助教。2015年より慶應義塾大学理工学部管理工学科専任講師(有期)。現在に至る。博士(工学)。セマンティック Web とオントロジーに関する研究に従事。情報システム学会、ACM、日本データベース学会、電子情報通信学会の各会員。



福田 直樹 (正会員)

1997年名古屋工業大学工学部知能情報システム学科卒業。1999年同大学院工学研究科電気情報工学専攻博士前期課程修了。2002年同大学院博士後期課程修了。同年、静岡大学情報学部情報科学科助手。2007年同助教。2010年同講師。2015年より同大学院情報学領域准教授。現在に至る。博士(工学)。2012年山下記念研究賞。2015年3月米国ハーバード大学短期訪問研究員。IEEE-CS、ACM、日本ソフトウェア科学会、情報システム学会各会員。情報処理学会シニア会員。2011～12年まで情報処理学会論文誌ジャーナル・JIP 編集委員会知能グループ副査。2013年同主査を歴任。