

学生海外調査研究	
The 10th International Semantic Web Conference に参加して	
大西 可奈子	
理学専攻	
期間	2011 年 10 月 21 日～2011 年 10 月 29 日
場所	ボン（ドイツ）
施設	Hotel Maritim

内容報告

1. はじめに

2011 年度「国際的に活躍する女性リーダー育成」推進事業の支援を受け、2011 年 10 月 23 日から 27 日までドイツのボンで開催された The 10th International Semantic Web Conference に参加してきたことについて報告いたします。残念ながら本会議およびワークショップで発表をすることはできませんでしたが、初めての ISWC2011 の参加は私にとって非常に大きな経験となりました。一つ目は、セマンティック Web 研究のトップの学会に参加することが、これからの研究意欲に結び付くことです。二つ目は、会議で発表される研究内容を知ることにより、現在行っている研究をより良いものすることができることです。

まず一つ目ですが、先ほど述べましたように、この会議は私が研究しているセマンティック Web 研究においてトップの国際会議です。私は今の研究分野にしたのが比較的最近であったこともあり、ISWC の存在は知っていたものの、参加することはできていませんでした。また ISWC は——ISWC に限らず、トップクラスの国際会議は——参加費も非常に高価で、かつ欧州で開催されることが多いため、気軽に参加できない会議でありました。ですので、今回、このように補助をいただくことができ参加できたことを本当に嬉しく思うと同時に感謝しております。

今回、会議が開催されたドイツは Semantic Web に関する研究が世界的にみても盛んな国です。そのためか、ISWC2011 には、今年も Massachusetts Institute of Technology の Alex Pentland 教授をはじめとする著名な先生方が基調講演にいられました。また、日本からは発表こそなかったものの、大阪大学の古崎先生、国立情報学研究所の武田先生等日本における著名な Semantic Web 研究の先生方の参加も見られ、お話をさせていただくこともできました。また、今回は先日発表されたばかりの、日本で行われる Semantic Web チャレンジの宣伝もされると聞いております。Semantic Web チャレンジは ISWC2011 のプログラムにもありますように、主に欧州においてすでに行われているチャレンジであります。内容は、ここでは具体的には述べませんが、例えば Semantic Web の技術のひとつである Linked Data の作成コンテスト等がその一つです。日本においては Semantic Web の研究はまだ遅れがちですが、日本でも Semantic Web チャレンジが行われるのだということをアピールすることにより日本の Semantic Web 研究を世界にアピールするチャンスだと考えられます。

二つ目に関しては、通常関連研究を調査する場合、ひとつの学会に発表されている論文をすべて調査するという事は行いませんし、またやろうと思っても困難な作業です。関連研究の調査は文字通り、自分の研究に深く関連する研究を調査します。しかしそうして研究を進めていくうちに、しばしば視野が狭くなりがちです。そのため、今回のような、ある分野で最先端の研究が発表される場で、それらの発表をすべて聴くということは、その研究分野の現在の流れを知るだけでなく、自分の研究を客観的に見直す非常に良い機会でもあります。

以上の二点が、今回の学会参加によって私が得た大きな経験です。

2. Consuming Linked Data 2011 について

ISWC2011 のプログラムにおいて、まず最も興味深く、かつ私が現在行っている研究と近い Consuming Linked Data 2011 ワークショップについて述べたいと思います。本ワークショップには私も論文を投稿させていただきましたが、力及ばずアクセプトされることができませんでした。従って、本ワークショップに参加することにより、私の研究の弱点や問題点を洗い出せることを期待し、参加

させていただきました。

COLD2011 ではフルリサーチが 16, システムが 3, ビジョンが 6 の発表がありました。

まず The University of Texas at Austin の Daniel P. Miranker 教授から, Removing the Potholes in the Road to the Semantic Web と題したセマンティック Web が広がっていくために除去しなければならない問題についての基調講演がありました。私が研究している Semantic Web は, 今後の Web となるのが前提となっていて行われている研究ですが, 実際のところ Semantic Web が広まるためには様々な障害があると言われています。この講演でより具体的にその問題を理解することができたように思います。具体的には, より大きな Linked Data のデベロッパーコミュニティを形成すること, コンテンツ/アプリケーションの作成ツールが必要であること, かつ簡単で, 親しみやすいものであるということ, 等が教授によって挙げられました。これらの問題は様々な形で研究されており, 今後より良い解決をなされることが期待されています。

続いて, 研究発表について簡単に述べていきたいと思います。

Fouad Zablit, Mathieu D'Aquin, Stuart Brown, Liam Green-Hughes による Consuming Linked Data within a Large Educational Organization では, イギリスの大学において, 実際に Linked Data が活用されている例を 3 つのユースケースにより, 技術的なものを含めて紹介されました。イギリスは Semantic Web 研究だけでなく, こうした実際の利用が非常に盛んな国であり (BBC 等も利用しています) またこういった事例が出たのかと感慨深く思いました。

Elena Simperl, Barry Norton, Denny Vrandečić による Crowdsourcing Tasks within Linked Data Management では, Linked Data やリンクド・サービス・テクノロジーを改良することによって人間とコンピュータの知性が相互に存在できるようなフレームワークを提供するためのアイデア等が述べられました。これまでは確かに, Linked Data を一定のクオリティに保つにはどうしても人間の手によるものが必要でしたが, この研究が示すように, 今後は人間とコンピュータの知性が相互に存在するためのフレームワークが必要になってくるのではないかと感じました。

Xingjian Zhang, Jeff Heflin による Using Tag Clouds to Quickly Discover Patterns in Linked Data Sets では, RDF データのためのタグクラウド技術について述べられました。セマンティック Web においてはタグクラウド技術の研究が非常に盛んですが, まだ新しい研究がなされているということが非常に勉強になりました。本ワークショップにおいても, これだけでなく他にも数件タグクラウドに言及している研究がありました。

Magnus Stühr, Dumitru Roman, David Norheim による LODWheel - JavaScript-based Visualization of RDF Data では, RDF データ可視化のためにこれまでに作られた JavaScript ライブラリの評価を行うと共に, JavaScript ベースの RDF データ可視化するための新しい技術 LODWheel を提案されました。グラフ上の可視化はカラーで直観的にわかりやすいと感じました。また, ノルウェーで行われている二つのプロジェクトも紹介されました。

Khadija Elbedweihy, Suvodeep Mazumdar, Amparo E. Cano, Stuart N. Wrigley, Fabio Ciravegna による Identifying Information Needs by Modelling Collective Query Patterns では, クエリログ分析, およびどのようにそのような分析を利用するかを形式化するためのアプローチが示されました。さらに SEMLEX という, 質問の分析によってユーザの情報ニーズの調査を促進するビジュアル化されたインタフェースも提案されました。また, DBpedia を利用したユースケースによる説明もありました。DBpedia は詳しくは述べませんが, 私も研究で使わせていただいております。DBpedia はこの研究以外にも多くの研究において利用されており, 私も今後も利用させてもらおうと思いました。

Daniel M. Herzig, Thanh Tran による One Query to Bind Them All では, 複数のデータセットにひとつのクエリを投げることににより全ての結果をもらうためのアプローチを 5 ステップで提案されました。また, 先の研究にも出てきた DBpedia 及び IMdb を使った実験も公開されました。私は IMdb (映画についての情報の RDF バージョン) のことを知らなかったもので, それを知れたことも収穫となりました。IMdb は今後, 研究で使わせてもらおうかもしれません。

Florian Schmedding による Incremental SPARQL Evaluation for Query Answering on Linked Data では, RDF データのための標準照会言語 SPARQL を Linked Data へ適応する時に生じるクエリ過不足問題に対して, SPARQL 代数の形式的分析によって解決を試みています。これは, 挿入と削除の数に依存する推測上最良である計算を選ぶオプティマイザのデザインのための増分の評価のコストを評価することができます。さらに, Linked Data から SPARQL データセットの構築も提案されました。本研究を知るまで, 私は研究を行う上で自然に SPARQL を Linked Data に利用していましたが, ここにも実は隠れた問題があったのだということに気づきました。SPARQL そのものの問題に研究として寄与する予定は今のところありませんが, このような問題もあったことを知れたという意味で非常に勉強になりました。

OlafGorlitz と Steffen Staab による SPLENDID: SPARQL Endpoint Federation Exploiting VOID Descriptions では、RDF データも従来のデータベース同様、利用者が意図しなくてもスムーズに利用できるようになる必要があるという考えに対し、Linked Data と従来のデータベースのアプローチには大きな差があるため、RDF に従来のデータベース技術を適応することは簡単ではないと問題提起がなされました。そこで、SPLENDID (voID descriptions から得られた統計データに基づいた SPARQL エンドポイントを連合させるための質問最適化戦略) が提供されました。

Andreas Schultz, Andrea Matteini, Robert Isele, Christian Bizer, Christian Becker による LDIF - Linked Data Integration Framework では、LDIF (Linked Data Integration Framework) フレームワークのアーキテクチャーについて記述され、ライフサイエンス・ユースケースのパフォーマンス評価も示されました。LDIF は、Web の異種混合データをデータの出所を保持したローカルの目的表現に翻訳するための、Linked Data アプリケーションにおいて利用できるフレームワークです。フレームワークそのものを利用するかどうかはわかりませんが、Linked Data アプリケーションを作成している私にとって、本研究のフレームワークの考え方そのものが非常に興味深いものでした。

Peter Haase, Michael Schmidt, Andreas Schwarte による The Information Workbench as a Self-Service Platform for Linked Data Applications では、Linked Data アプリケーションを作成するのにサポートするインフォメーション・ワークベンチが提案されました。Linked Data アプリケーションの発表があることを期待していたのですが、本研究や前述した LDIF のように Linked Data アプリケーションを構築するためのフレームワーク研究も盛んなのだということを知れてよかったです。

Valentina Presutti, Lora Aroyo, Alessandro Adamou, Balthasar Schopman, Aldo Gangemi, Guus Schreiber による Extracting Core Knowledge from Linked Data では、任意のデータセットの中心となる type や property を自動で取り出す手法について語られました。データセットの知識はしばしば不定形にて組織されているため、一般的にそれに対してどのようなクエリを書けばいいのかわかりません。そこで、知識パターンに基づいたデータセット自動解析アプローチが提案されました。この研究は非常に興味深かったです。私も Linked Data のあるデータセットを使おうとした時、しばしばこの問題にぶち当たるからです。現在のところ自分で調べるしかないのですが、これが自動で正確になされることができるのであればすごいことだと思います。また、私もデータセットからある知識を抽出する研究を行っているため、目的は違いますが手法や考え方は参考になることもあるのではないかと思います。これについては、後日、熟慮してみるつもりです。

Benjamin Zepilko, Brigitte Mathiak による Defining and Executing Assessment Tests on Linked Data for Statistical Analysis では、統計的な解析のための Linked Data の Executing Assessment テストが定義されました。これに限ったことではありませんが、本ワークショップにおいては Linked Data に統計的处理を行う扱う研究は多かったように感じました。

Marcus Cobden, Jennifer Black, Nicholas Gibbins, Les Carr, Nigel Shadbolt による A Research Agenda for Linked Closed Data では、Linked Open Data—その名の通り、広く公開された Linked Data —と反対に、クローズドな Linked Data である Linked Closed Data が提案されました。これは vision paper (先見性のある論文) カテゴリで発表されました。確かに Linked Open Data はしばしば法的・プライバシー的に問題があり、将来的に Linked Closed Data の定義が必要となってくるかもしれないと思っていたので、とても興味深い発表でした。

Julia Hoxha, Anisa Rula, Basil Ell による Towards Green Linked Data では、作られた Linked Data がグリーンかどうかを判断する条件を策定し、それをチェックするシステムが作成されました。発表にもあったように、近年 Linked Data は膨大な数になってきているけれども、そのクオリティ自体は下がってきていると言えます。クオリティの低いコンテンツが集まって、大きなサイズのデータになっている場合もあります。そのようなミスが含まれている Linked Data を減らすために、このシステムはある程度有効なのではないかと感じました。私は現在のところ Linked Data の応用研究を行っているため今使うことはなさそうですが、将来 Linked Data の制作をすることがあれば利用してみたいと思います。

COLD2011 の Closing Discussion では、昨年と今年のジャンルのバラつきについて言及されました。昨年は、

- Meta-data based consumption 1
- Dataset dynamics 1
- Knowledge discovery 2
- User interaction, usability, visualization 1
- In use and systems 1
- Temporal Linked Data 1

- Services and Linked Data 1
- とバラつきがあったのに対し、今年は、
- Query processing over multiple datasets 3
- Meta-data based consumption 1
- Knowledge discovery 3
- Information quality/trustworthiness 1
- User interaction, usability, visualization 2
- In use and systems 4

と、比較的偏った結果となりました。しかし、これに関しては悲観的に考えず、来年の投稿を待つと言った意見が多く見られました。また中には、もっと使えるアプリケーションとしての **Linked Data** に関する論文が増えて欲しいという意見もありました。私は応用分野をやっているの、こういった考え方がすることは非常に好ましいことです。

最後に、今年で二度目を迎える **ISWC 2011 Linked Data-a-thon** において、ノミネートされた **Linked Data** を使ったアプリケーション 4 つの内から投票でトップを決めるためのデモが行われました(実際にデモされたのは以下の 3 つでした)。投票はインターネットで行われました。

- ISWC 2011 Conference Explorer
- Conformation
- Friends and Facets

どれも非常に素晴らしい出来だったのですが、**Linked Data** の新しい使い方という意味ではどれも少しインパクトが弱かったように思いました。そのあたりは **Linked Data** を使うに当たって、今後の課題なのかもしれません。

COLD2011 全体を通して、採択された論文は、デモが行えるような完成度の高い可視化(タグクラウド含む)や分析が特に多かったように感じました。また、かなり大規模な研究も多く、大きなプロジェクトがたくさん動いているのだということを痛感しました。このような中に、大学の研究室の研究を採録させるためには、大きなプロジェクトではできないような一風変わったアイデアやアプローチが必要なのではないかと感じました。また、**Linked Data** を利用した応用研究はあまり多くなく、その点においては少し期待外れでもありました。丁寧に造られたソフトウェアの紹介は幾つかあったのですが、それが本当の意味で使え、使ってもらえるアプリケーションなのかということは疑問でもあります。また、アプリケーション作成のためのフレームワークの提案も複数ありましたが、今後はそのようなフレームワークを利用して、**Linked Data** を利用するからこそ便利であり、かつユーザに実際に利用されるアプリケーションを提案することが重要なのではないかと感じました。しかしながら、今後の私の **Linked Data** の研究の方針を考える上で、本ワークショップへの参加は多くの収穫があったように思います。

3. ISWC2011 メインカンファレンスについて

2011 年は素晴らしいサービス (facebook や twitter 等) がたくさん盛り上がった年でした、というコメントから始まったオープニングトークは、早くも来年の **ISWC2012** への意気込みと今回のベストペーパーにノミネートされた論文等が発表されました。また、メインカンファレンス一日目には MIT のプロフェッサー Alex Pentland 氏による基調講演「Building A Nervous System for Society」も行われました。この基調講演はタイトルの通り、アクティブコントロールネットワークのため健康や運送手段やエネルギーなどの社会システムを再構築するためにはどうしたらいいのか、について「普及力のあるセンシング」「人の行動のモデリング」「活発な基盤 (インフラ)」という三つの観点から論じられました。人の振る舞いをデモするのは非常に大変ですが、現在ならば例えばスマートフォンなどのセンサーからそれぞれの状況を把握してやる方法などが考えられます。人が何かを決定することに対しては、多くのネットワークが影響しており、単純ではありません。それをいかにして構築するかは、確かに難しい問題であると思うと同時に、**Semantic Web** 技術によって解決できる問題なのかもしれないとも思いました。また、個人情報は今やインターネット上の石油である、という言葉が印象的でもありました。

その他、メインカンファレンスでの発表について簡単に述べたいと思います。

Mike Dean がセッションチェアを務めた **In-Use Ontologies and Linked Data** において、興味深い四つの報告がありました。そのうち、二つについて詳細を述べます。Christian Seitz and Rene Schonfelder による Rule-based OWL Reasoning for specific Embedded Devices, そして Peter Haglich, Robert Grimshaw, Steven Wilder, Marian Nodine and Bryan Lyles による Cyber Scientific Test Language の後に発表があった, Xing Niu, Xinruo Sun, Haofen Wang, Shu Rong, Guilin Qi and Yong Yu

による Zhishi.me - Weaving Chinese Linking Open Data では、現在 Linked Open Data において中国語でかかれている知識は乏しいため、中国の Linked Data を、中国語で作ることによる問題を解決しながら作成したことについて述べられた。確かに漢字は Semantic Web には少々不向きな部分があるが、それをクリアしたうえで、Lookup サービスや SPARQL エンドポイントも完備した Linked Data が作成されていました。Linked Data にするためのオリジナルデータソースは Baidu, Hudong, C-Wikipedia からとってきたそうです。また、最後の発表である、Miriam Fernandez, Mathieu d'Aquin and Enrico Motta による Linking Data Across Universities: an integrated video lectures dataset では、機関を横切って教材を統合し共有する必要があることから、Linked Data 技術を利用した大学間をつなぐ教育関係の情報のリンクを作成すると共に、ビデオのデータセットも作成したそうです。実際に使われている例を見せつつ説明されました。質疑応答では、他にオープン大学における似た取り組みがあることも言及されました。

また、Aldo Gangemi がセッションチェアを務めた RDF Data Analysis でも、Andreas Thor, Philip Anderson, Louisa Raschid, Saket Navlakha, Barna Saha, Samir Khuller and Xiao-Ning Zhang による Link Prediction for Annotation Graphs using Graph Summarization や Harris Lin, Neeraj Koul and Vasant Honavar による Learning Relational Bayesian Classifiers from RDF Data, Yang Yu and Jeff Heflin による Extending Functional Dependency to Detect Abnormal Data in RDF Graphs, Evgeny Kharlamov and Dmitriy Zheleznyakov による Capturing Instance Level Ontology Evolution for DL-Lite といった興味深い発表が多かったです。

その他、三日間に及ぶメインカンファレンス期間には、たくさんのセッションで多くの興味深い研究発表を聴くことができました。

4. おわりに

はじめにも述べましたように、今回の経験を通して私は大きく分けて二つの収穫がありました。一つは、トップカンファレンスに参加している同じ分野の研究者との交流を通して、国際的に活躍できる研究者を目指したいという意識の向上です。もう一つは、該当分野の研究の大きな流れを、一步引いた場所から見ることにより、今後の研究方針をより良いものへとする知識の獲得です。私はこれまで、本学会でも多く発表されている Linked Data という技術を使った応用研究をし、国内での発表を行ってきました。今回の学会では、私がやっているような応用研究として、Linked Data を利用したアプリケーションの紹介などが複数あり、今後の研究の進め方を考える上で非常に参考になりました。具体的には、博士論文執筆において二章の関連研究にて本学会で発表された論文を参照させていただく予定である他に、今後の研究に技術的な部分でも参考にさせていただく予定です。

また、本学会で発表された研究成果は、国家ぐるみの大きなプロジェクトの一部、あるいは集大成であるものが非常に多かったように思いました。そのような発表者の中には、研究プロジェクトのリーダーとして関わっているたくさんの女性研究者がいました。これまでの私の国際学会での発表は二回と決して多くはありませんが、今回の経験を通して彼女らのような国際的に活躍できる女性研究者になりたいという意識が強くなりました。

今後は、今回得た知識を活かして研究を進めると共に、博士論文を執筆し、国際的に活躍できる女性研究者を目指して邁進していきたいと思います。

おおにし かなこ／お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻

指導教員によるコメント

International Semantic Web Conference はインターネットをより高度に進化させるという主旨をもったトップの国際会議であり、今回、大西可奈子氏が「学生海外派遣」プログラムを利用させていただき、参加させていただいたことには大変大きな意味がありました。彼女の研究は、現在、Semantic Web 技術の枠組みにおいて中心的な技術となっている Linked Open Data を利用した新しい情報処理技術を提案するということです。この研究を進めるにおいては世界のトップの動向を絶えず把握することが必要であり、それは情報を Web で収集することだけではなく、その分野の多くの研究者が一同に会する国際会議などに参加し、短期間に集中して著名な研究者らと情報交換をすることが大変重要となります。現在、彼女は博士後期課程 3 年生であり、博士論文の作成をおこなっています。今回の会議参加において、現在の最先端の研究に触れることができ、そのことは彼女の博士論文作成にも大きな意味をもったものになったと思います。また、彼女はこれから女性研究者として研究を続けていくことになっておりますが、世界で女性研究者がどのように一流の研究を行っているかを見る機会もあったようで、彼女の研究者としての意識の向上に大変役立つことができたと思っております。こ

のような機会を大西可奈子氏に与えてくださったことに深く感謝の意を表します。
(お茶の水女子大学大学院人間文化創成科学研究科（自然・応用科学系）・小林一郎)