プロジェクト名：　異分野研究資源共有・協働基盤の構築
（略称：サイエンス 3.0 基盤構築）

プロジェクトディレクター：　新井 紀子教授（国立情報学研究所）

〔1〕 研究計画・研究内容について

(1) 目的・目標

自然科学から人文科学にわたる異分野の「知」と「人」の共有・連携を行い、情報や研究人材の効果的な活用や研究協力・共同研究の促進を行う学術知共有・学術連携促進基盤を構築し、実用に供する。その手段として、まず、インターネット上で様々なところに散在する学術情報および研究支援サービスを結合して利用可能とするプラットフォームを構築する。このように収集された学術データを研究対象として新しい検索技術・機械学習・データマイニング・ユーザインタフェイス技術・可視化技術等の研究開発を通じて、研究者あるいは研究分野・研究プロジェクトごとにパーソナライズされた学術情報・学術サービスの提供を目指す。

具体的には、サブテーマ「研究資源に関する情報推薦基盤の構築」においては、機械学習・データマイニング・オントロジー関する研究を通じて、情報推薦に関して世界をリードする独自技術を開発する。サブテーマ「学術リソースのためのオープン・ソーシャル・セマンティック Web 基盤の構築」において、セマンティックウェブ技術およびデータベース連携の研究開発を通じて、研究者向け次世代ウェブサービスの構造に関する技術開発を行い、散在する学術研究資料が有効活用するための基盤を整える。サブテーマ「多様な知的情報源を結合・融合・再構成する連想情報処理基盤の構築」において、論文情報や書誌情報といった定型的なデータ以外にも、発表資料、コースウェア、研究データなどの異種データをリンケージした上で高速な連想検索を行うための技術の確立を目指す。以上のサブテーマによる、研究開発をサブテーマ「融合研究を加速するための情報共有クラウドサービスの確立」で統合し、世界をリードする次世代研究者サービスを構築し、日本の学術知共有・学術連携を促進することを目指す。

(2) 必要性・重要性（緊急性）

インターネットを通じて様々な学術情報・学術サービスが公開・提供されるようになったが、単に Web に公開しただけは相互運用性がなく、情報を十分に活用することはできない。特に、近年学術分野においても情報爆発が起こっており、これに対応するため、学術情報に関する各種電子アーカイブが整備されつつある。また、多種多様な分野における研究人材データや研究用のデータベースも電子化されてきた。世界的な研究開発の加速・競争の激化の中、整備されつつある研究データ・論文アーカイブ・人材データベース・研究用ミドルウェア等をいかに有機的に連携し、柔軟かつ機動的に共同研究を進めるかということが、日本が科学立国としての地位を維持する上で、鍵となる。しかしながら、現状においては、これらの学術情報・学術サービスを有機的に結合する手段は未成熟であり、人材と研究に関する連携力が充分に発揮されているとはいえない。また、情報技術から遠い学問分野においては、このような潮流の認識が諸外国に比べて進んでおらず、取り残される危険性がある。この問題を解決する手段として、すべての学問分野の研究者にとって使いやすくまた柔軟性のある学術知共有・学術連携促進基盤を構築する必要性がある。

(3) 期待される成果等（学問的効果、社会的効果、改善効果等）

既存の大規模データベースを有機的に結合するための「ハブ」となるシステムを研究者に提供することにより、学術知共有・学術連携が促進される。特に、異分野での連携促進が期待できる。また、本シ

ステムを実運用システムとして全国の研究者に提供することにより、日本最大級の「生きた」学術データベースが構築され自律的に増殖していくことになる。このことにより、主として 3 つの社会的波及効果がある。第一に、本システムに蓄積されたデータを研究対象として新しい検索技術・データマイニング・情報推薦・ユーザインタフェイス技術・可視化技術等の開発が進むことが期待できる。第二に、本システムをサービスとして利用する研究者は、多様かつ膨大な学術データベースから、自分の研究分野や研究関心にあわせた最適な研究情報が「推薦」され、編集された上でタイムリーに届けられる。また、研究を支援するような各種サービスが、クラウド基盤を通じて提供される。これは、競争が激化している各研究分野において、日本の研究者が国際的優位性を勝ち取る上で、たいへん重要である。第三に、本システムに蓄積された研究情報が国民に随時公開されることにより、多様かつ信頼がおける科学コミュニケーションの場が副次的に実現されることである。

(4) 独創性・新規性等

本プロジェクトでは、多様な異種学術データを大規模に収集した上で、情報および統計の技術を駆使し、各研究者に対して、パーソナライズされた情報およびサービスを提供するという極めて先進的な取り組みを行う。本分野は、1－5 でも説明するとおり、世界中の研究機関・研究者向け商用サービスが重要視し、取り組みを本格化させているところでもある。その中で、本プロジェクトは以下の点において、優位性および独創性がある。

まず、国立情報学研究所は国内有数な学術データベースを有しており、また、情報・システム研究機構の融合研究センターはライフサイエンス統合データベースを有している。大学共同利用機関法人として、各種の機関リポジトリやデータベースとの連携関係も深い。これらデータベースと結合することで、他機関では到底実現不可能な大規模な情報流通基盤が実現可能となる。本プロジェクトが具体的に実現される基盤である NetCommons は 2007 年には国際学会 IASTED 主催第 3 会国際ソフトウェア競技会で最優秀賞に選ばれたほか、2009 年には IPA より日本 OSS 奨励賞を受賞するなど国際的評価も高い。その上で、世界最速の連想計算エンジン GETA によるコンテンツ・コンパイル技術を用い、蓄積された情報源の特徴を計算機構として抽出する。さらに情報源同士の相互作用に活用し、研究者の特性をデータマイニング技術によって抽出した上で、パーソナライズされた情報推薦を行うことは、非常に先進的・独創的な取り組みである。また、単に先進的・独創的な研究であるだけでなく、研究開発成果が直ちに、産学官を超えたすべての日本人研究者に提供される。その意味でも、社会貢献の度合い、費用対効果も極めて高い。

(5) これまでの取り組み内容の概要及び実績

本研究に先立って、第一期新領域融合研究「分野横断型融合研究のための情報空間・情報基盤の構築」においては、融合研究を加速するためのバーチャルラボシステム NetCommons を構築し、オープンソースソフトウェアとして公開している。また、異種情報の結合・分類手法に関しての研究を進め、世界最速の連想計算エンジン GETA によるコンテンツ・コンパイル技術を確立して、異なる情報源同士の相互作用を情報探索に利用する想・IMAGINE システムを開発した。さらに、大規模リンケージ情報の研究では、国立情報学研究所で公開中の「科学研究費補助金データベース」を情報源として、約 13 万人の日本人研究者について統一的な研究者 ID 番号の情報を提供する「研究者情報サーバ」プロトタイプ版システムを拡張し、他のデータベースとの統合のための機能整備を行った。これらの成果を概念レベルだけでなく、具体的に融合させ、平成 20 年度には、「状況に埋め込まれた人間の相貌をデジタルに表現する技術の研究」において、NetCommons を基盤として、コンテクスト（状況）の中で、さまざまな相貌をみせる人間の活動にフィットするポストウェブの技術の開発を目指し、今回提案するサイエン

ス 3.0 基盤のプロトタイプとなる Researchmap α 版の開発を行った。具体的には、多様な学術情報データベースから、研究者 ID をキーとして論文情報・研究者経歴等の学術情報を複数のデータベースから自動取得する方法を開発し、研究者の CV データとして編集・公開する機能を実装した上で（担当：相澤、大向、新井)、CV データを軸として、興味関心の近い研究者を分野横断的に検索する技術を開発し(担当：新井・高野・丸川・舛川)、研究者の研究コミュニティの形成および運営を支援するための基盤サービスの提供を試行し、既に 1300 人を超える研究者が実際に試用している。今後も利用者が増加することが見込まれ、より多くの研究データが蓄積することが確実となっており、次期新領域融合研究を開始する準備が整っている。

(6) 国内外における関連分野の学術研究の動向

海外の学術機関の動向については､フィンランドが健康バイオ分野でセマンティック Web 技術を利用した広範なデータベース連携を実現している。しかし主たるターゲットは公共的機関がもつデータであり、研究データなどはあまり対象となっていない。また EU では Europeana プロジェクトが各国の博物館データの統合を進めているが、統合の程度はあまり深くない。

商用サービスを含めた動向としては、研究者が独自の ID を取得できる Researcher ID というサービスを Thomson 社が開始し、また、研究者の情報発信支援を Academa。edu が提供するなど、研究者向けに学術情報サービスを提供する試みがまさに始まったばかりであり、世界的関心が非常に高い。しかし、これらのサービスは論文情報販売を目的とした情報収集および顧客囲い込みのためのサービスであり、学術情報を横断的に活用しながら共同研究を推進する基盤を目指しているわけではない。

〔2〕 研究計画

(1) 全体計画

学術情報は、かつてはきわめて狭く固定的な方法で流通していた。流通の範囲は自らの分野の専門家に限定され、方法も学術雑誌における論文といった出版に限られていた。しかし、本来、学術情報はもっと広く柔軟に流通すべきである。学術成果は単に結果を論文として発表するのではなく、利用したデータや結果に関するデータといった情報、研究過程といったものも公開・共有されることが、開かれた科学技術の発展上は望ましい。また学際的な研究も盛んになっている現在、自分の分野だけで利用可能な情報流通は適しているとはいえない。一方で、科学技術における発見や発明が、富の源泉であることは、科学技術の 4 千年を超える歴史の中で自明のことであり、研究過程を公開することは、研究者にとっても各国の科学技術戦略の上でも、慎重である必要がある。

ここに、研究者最新の学術研究データに 1 秒でも早くアクセスした上で、自らの研究成果および過程は、適切な共同研究者との間で安全に共有し、それを素早く商用化したり、研究成果として公知としたり、そのサイクルの中で、より大きな競争的資金やより良い共同研究者を獲得する、というニーズが、否が応でも高まる素地があるといえよう。学術研究データに関する多様なデータがデジタル化され、アーカイブされるようになった今、このことは一見、直ちに実現され得るかのように見える。しかしながら、そこにはいくつかの理論的・技術的な困難が存在する。

第一は、多様な学術研究データがウェブ空間上に爆発的に増加した結果、それらのデータにアクセスすることは概念的には可能であるが、現実には不可能に近い。そこで、研究者の知的生産活動にとって効果的で確実な検索技術が不可欠になる。ところが､研究者の在り方や興味関心分野は多種多様であり、必要とするデータも多種多様である。よって、ウェブ上に拡散する学術研究データが多様になればなるほど、個々の研究者に特化した形で、あたかも執事のように情報をリトリーブして的確に提供するためのプッシュ型の情報検索・情報推薦の技術が望まれる。ここに第二の困難がある。研究者の興味関心に

従って、ウェブ上の学術研究データの意味を発見・分類し、統計処理した上で、情報推薦することは、画像処理であればセマンティックギャップ、人工知能であればフレーム問題に相当する、セマンティックとシンタクスをつなぐ非常に困難な問題だからである。そこで、我々は、データマイニングとオントロジーを用いた手法と、ソーシャルメディア的手法を用いてユーザ自身からフィードバックを得る手法と、外部の信頼おけるデータとそれに付与された情報を活用した連想検索の手法を統合することで、この課題の克服を目指す。

<table>
<tr><td rowspan="2">テ ー マ</td><td>H22 年度</td><td>H23 年度</td><td>H24 年度</td><td>H25 年度</td><td>H26 年度</td><td>H27 年度</td></tr>
<tr><td>(予備研究)</td><td></td><td></td><td>中間評価</td><td></td><td>事業化</td></tr>
<tr><td>全　　体</td><td colspan="5">実システムへの適用・Web 空間との連携・実証研究・改良</td><td>事業化</td></tr>
<tr><td>サブテーマ１</td><td rowspan="3">準備調査研究
プロトシステムの開発</td><td colspan="2">「情報推薦」技術の研究開発</td><td colspan="2">「情報推薦」技術の改良と深化</td><td rowspan="3">他のシステムへの応用</td></tr>
<tr><td>サブテーマ２</td><td colspan="2">セマンティックウェブ技術の研究開発</td><td colspan="2">セマンティックウェブ技術の改良と深化</td></tr>
<tr><td>サブテーマ３</td><td colspan="2">多種データ間の連想検索技術の研究開発</td><td colspan="2">サブテーマ３はサブテーマ４に統合</td></tr>
<tr><td>サブテーマ４</td><td>連携準備</td><td colspan="2">国内学術分野における連携強化</td><td colspan="2">産業界・海外との連携強化</td><td>国際展開</td></tr>
</table>

(2) 各年度の計画

平成 24 年度（中間評価）

サブテーマ１では、研究者の論文検索における嗜好を調査し、利用要求を分析する。研究者個人プロファイルと論文との関連度指標について検討し、特に研究者の多様な利用要求に対応するための推薦手法の開発・実証を行う。専門用語辞書やウェブ情報などの外部情報源の活用方法を検討するとともに、言語的な手法に基づく論文記述の解析や記述どうしの相互の参照関係抽出の手法について課題を整理する。また、論文記述の深い解析を支える基盤として、構文解析技術の解析対象を、抄録などの平坦で整えられた文から、構造・表現に柔軟性のある論文全体へと拡張するための基本枠組を構築する。

サブテーマ２では、本年度までにデータ中心型研究の基盤のプロトタイプを完成させる。具体的には基盤的ソフトウエア開発および基盤的データベースを完成させる。また応用的ソフトウエアとして GIS を含む多様な LOD データを連係される。

1. 基盤ソフトウエア環境構築

1.1 データ統合ソフトウエア開発

複数のデータサイトからデータを収集すると、それらの関係の管理が重要になる。本プロジェクトではデータの多様性を維持したままデータを統合するために、共通で核となるデータとそれに結びつけられた個別のデータという構造でデータを管理する。このようなデータ管理を可能とする仕組みを考案して、実装を行う。

2. 基盤データベース構築

2.1 生物種情報 LOD

前年度から廃止した生物種情報 LOD のスケールを拡大して、一通りの生物種を格納し、和名を含む名称検索やタクソン構造が辿れるようにする。また EOL や GBIF など国際的に重要なデータサイトとの連携を実現する。

2.2　シソーラス・百科事典・辞典情報

分野横断的情報としてはシソーラス、百科事典や辞書の情報はハブとして機能する。百科事典としては日本語 Wikipedia を対象として、その Linked Data 化（日本語 DBpedia）を行う。辞書情報としては日本語 WordNet などのオンライン辞典を統合して Linked Data 化する。また関係する学術分野のシソーラスの Linked Data 化も行う。

3.　研究コミュニティ支援サービス構築

3.1　環境プロジェクト連携

前年度に引き続き、国立遺伝学研究所の GBIF（Global Biodiversity Information Facility）活動と連携する。上記の生物種メタデータデータベースを基盤して使い、GBIF データの取り込み支援システム、データの連係支援システム、可視化システムなどを実装する（共同研究者：神保宇嗣（国立科学博物館））。

3.2　GIS プロジェクト連携

前年度に引き続き、国立極地研究所と連携して、GIS とのデータ統合の仕組みを研究する。（共同研究者：小林悟志（極地研））。

3.3　データベース検索支援

上記での構築した生物種メタデータベースを用いて、DBCLS が収集したデータベース検索およびデータ検索における検索支援を行う。ユーザが入力した語を生物種メタデータベースに問い合わせて関連するタクソンの追加やその和名学名変換を行うことでユーザが必ずしも適切な語を入力しなくても検索できるようにする。（共同研究者：川本祥子（DBLCS））

サブテーマ３では、震災アーカイブ内の写真や国立極地研究所の観測データなど、情報が取得された日時と場所だけが記録されたコンテンツが大規模に生成されている。これらのコンテンツについて日時や場所を指定しての検索は可能であるが、基本的には写真データ内、場所データ内のように同一種類のコンテンツ内で閉じており、今まで培われている大量のテキストコンテンツと同じ枠組みで検索することができない。そのため時空間コンテンツを格納したデータベースとテキスト情報からなるデータベースとを横断的に検索する仕組みが求められている。

連想検索は、単語ベクトル化された文書集合に対し、自然文や複数文書からの検索機能を提供するものである。複数のデータベースに対して横断検索するときには、どのデータカラムを検索対象として指定するかなど、それぞれのデータ構造を把握しお互いに連携する必要があるが、連想検索ではおのおののデータベースが自身で文書･単語行列を構築しておけば容易に横断検索できる仕組みを持つ。

画像や映像などのテキストベースでないコンテンツでも、タイトル、作者、説明文などのテキストから構成されるメタ情報が付与されていれば、それらの情報を利用して連想検索が可能である。

本研究では、テキストベースのコンテンツに時間･空間情報を自動的に付与し、メタ情報にテキスト情報を持たないコンテンツと横断的に連想検索ができる仕組みを提案する。本研究は主に２つの要素技術から構成される。

1.　テキスト情報から時間、空間の情報を抽出する

この技術は自然言語処理分野の固有表現抽出に該当する。ただし、固有表現抽出はテキスト中に明示的に出現している時間、空間を表現する文字列だけを抽出するのに対し、提案手法では、時空間情報を表さない固有表現に対して Wikipedia などのデータリソースの情報を用いて、時空間情報を付与する仕組みを開発する。

2.　連想検索に時空間の概念を取り込む

従来の連想検索では、単語から構成される語空間を検索対象としていたが、ここに時間、空間を取り入れる。複数のコンテンツを検索クエリーとした場合に対応するため、時間および空間の情報を確率分布の和で表現し、この値を離散化したベクトルを連想検索に用いる。連想検索時には、どの空間を重視

して検索をするかをユーザが指定できるようにし、それぞれの空間の寄与度合いを調節可能とする。最終的には国立極地研究所と連携して GIS 情報を核とするデータと国立情報学研究所の持つ論文や書籍の情報とを横断的に検索できる仕組みを開発する。

サブテーマ４では、前年度検討した API を実装し、他のデータベースへの提供を開始する。特に、府省共通研究開発管理システム（e-Rad）および、各大学の研究者総覧への提供を行う。これにより、研究者情報の循環が促され、異分野の研究者の融合が促進されることが期待できる。これまで研究開発された要素技術を改良した上で、Researchmap 上に統合し、ユーザからの評価の他、アクセシビリティ等に関して、外部の評価を受ける。

平成 25 年度

サブテーマ１では、論文抄録や全文データの言語的な解析による拡張について検討する。論文が扱う「手法」や「応用分野」などに関する情報を抽出して、推薦対象を論文から研究資源に拡大した推薦システムの実現を目指す。また、前年度に構築した論文全文推薦システムを論文の全文データと組み合わせることにより、より深い論文から実現される、推薦システムの更なる拡充を目指す。

サブテーマ２では、本年度からデータ中心型研究基盤の展開を行う。基盤システムとしての機能強化を図る共に応用的システムを作り、ケーススタディを進める。

まず、異なるデータサイトからくるインスタンス情報（個物に関する情報）は同じインスタンスを指していることがある。このインスタンスマッピングを効率的に行うプログラムを開発する。プロジェクトメンバーが開発したアルゴリズムを発展させ、実問題で適用できるようにし、実際にインスタンスマッピングを行い、データ統合を自動化する。GBIF データといった研究データにおける実データを用いる。

次に、構築された統合的データベースを可視化するためにいくつかのアプリケーションを作成する。たとえば、Linked Data アプローチで多様なデータが結ぶつくことを示すため、地図と地名から様々な情報にアクセスできるアプリケーションを作成する。地理・地名情報は多くの分野に共通するので、このアプリケーションを通じて様々な分野の情報、データが横断的に利用できる。地理情報であるので、PC 版だけでなく、携帯できるモバイル版も開発する。

以上で開発してきたさまざまなプログラム、システムおよびデータベースを統合的に利用できる環境を構築する。このプラットフォームを用いて、アプリケーションがデータを取得したり投入したりできるようにする。環境プロジェクトや GIS プロジェクトのシステム・データを統合する。

サブテーマ３（連想情報処理基盤の研究）については、中間評価に基づく PD の判断により、 独立のサブテーマとしては研究を中止して、必要に応じてサブテーマ１やサブテーマ４を推進する中で要素技術の活用を検討する。

サブテーマ４では、これまで蓄積したデータを基に、他のサブテーマ３で研究開発した要素技術を用いて異種データリンケージによる研究情報可視化のプロトタイプを実装する。また、サブテーマ１の要素技術を実装するために、研究論文・講演資料等を Researchmap 上で研究者が蓄積するための WEKO ベースの大規模 OpenDepo とその上の全文検索を検討・構築する。

平成 26 年度

サブテーマ１では、論文から抽出した情報を他のデータベースやウェブ上の情報に結びつけるとともに、API を経由して、外部の学術コンテンツサービス関連サーバと連携して、推薦を通して研究者どうしの協働を促進するための基盤システムを開発する。また、より広範囲な情報リソースとの結びつけを実現するべく、多様な記述スタイルを持ったテキストの解析を実現可能とする基盤技術の実現を目指す

とともに、論文から抽出した情報とその他リソースから得られる情報との質的な差異を摺り合せる手法について検討する。

サブテーマ２では、本年度からデータ中心型研究基盤の展開を行う。基盤システムとしての機能強化を図る共に応用的システムを作り、ケーススタディを進める。

まず、異なるデータサイトからくるインスタンス情報（個物に関する情報）は同じインスタンスを指していることがある。このインスタンスマッピングを効率的に行うプログラムを開発する。プロジェクトメンバーが開発したアルゴリズムを発展させ、実問題で適用できるようにし、実際にインスタンスマッピングを行い、データ統合を自動化する。GBIF データといった研究データにおける実データを用いる。次に、構築された統合的データベースを可視化するためにいくつかのアプリケーションを作成する。たとえば、Linked Data アプローチで多様なデータが結ぶつくことを示すため、地図と地名から様々な情報にアクセスできるアプリケーションを作成する。地理・地名情報は多くの分野に共通するので、このアプリケーションを通じて様々な分野の情報、データが横断的に利用できる。地理情報であるので、PC 版だけでなく、携帯できるモバイル版も開発する。以上で開発してきたさまざまなプログラム、システムおよびデータベースを統合的に利用できる環境を構築する。このプラットフォームを用いて、アプリケーションがデータを取得したり投入したりできるようにする。環境プロジェクトや GIS プロジェクトのシステム・データを統合する。

サブテーマ４では、平成 25 年度に構築した OpenDepo を研究者に対してリリースするとともに検索の高速化を図る。また、主要研究大学との Shibboleth 連携、API 連携を深める。これによって集約されたデータを基に、サブテーマ１およびサブテーマ２の実証実験を本格化させる。

平成 27 年度

サブテーマ１では、これまでに構築した基盤システムを活用するための実証システムを構築し、論文推薦手法および推薦の基本的アルゴリズムの評価を行うことによりその有効性を確認する。

サブテーマ２では、まず、論文に含まれるデータを抜き出し、データとして利用できるような発展的ソフトウェアの開発を行う。またデータの由来などの情報も同時に抽出する。この仕組みをつくることで論文とデータが同時に利用可能になり、データ中心型研究の新たな研究成果表現が発展することが期待される。さらに、学術分野ごとに存在する概念体系、専門用語体系を抽出してマッピングを行う。この学術オントロジーと論文、論文抽出データを同時に使うことで分野を超えた研究の理解とデータの利用が可能になる。

サブテーマ４では、サブテーマ２の研究成果を、統合プラットフォームのアプリケーションとして OpenDepo 上でデータ操作ができる環境を駆逐する。リポジトリに投入された情報からデータを取得し統合プラットフォームへデータを送ったり、データを入手できるようにする。リポジトリ自体がデータ中心型研究の環境として機能するようにする。また、これまで構築した、統合プラットフォーム、ResearchMap 統合、リポジトリ統合をシームレスにつなぎ、クラウド型データ中心型サービスの構築を構築する。

平成 28 年度以降の展開

これまでの研究成果の下、Researchmap は、異分野研究資源共有・協働基盤システムとして完成し、産業界や海外の共同研究者も含めて幅広い層の研究者に提供されるサイエンス 3.0 基盤として成果を上げていることが予定される。本システムを、事業として成立させることも可能であるが、ResearchersID 等の商用サービスが海外では成立していることから、民間への移転も可能であろう。どちらがより研究者のメリットになるかを検討した上で、自律的なサービスとしてテイクオフさせる方策を検討する。

また、大学機関リポジトリ、ライフサイエンス統合データベース、包括脳支援データベースなど各種データベースから CiNii や KAKEN に至る、研究情報サイクルの輪を完成させ、一度の入力によって、すべての情報が再利用可能な状態にし、利便性を向上させる。

各サブテーマの研究成果である要素技術については、オープンソースとして、あるいは、WebcatPlus、文化遺産オンライン等の学術情報サービスに反映させ、広く社会に還元する。

特に、サブテーマ 1 では、これまで構築した基盤システム、実証システムの評価を行うとともに、安定したサービス運用という形で研究成果を発信するための方策を検討する。

また、今後の研究はデータ中心型研究の方向に向かっていくことが予想される。サブテーマ 2 では、データを保管場所、保管方法などにとらわれずシームレスに利用可能し、データ操作が自由に行え、データを用いた研究がその場で実施し公開できる、データに基づくバーチャル研究環境が必要なると思われる。これまでの本研究の成果を生かしてこのような分野を超えてデータを自由に操作できる環境を構築する。

〔3〕 研究推進・実施体制

本研究の推進にあたっては、(1)既存学術データベースと連携するためのセマンティックウェブ技術 (2)異種データベースをつなぐリンケージ技術および検索技術 (3)大規模データマイニングおよびオントロジー技術が不可欠となる。そこで、本研究プロジェクトを 4 つのサブテーマに分け、3 つのサブテーマにおいて、(1)(2)(3)に関する研究開発を行った上で、4 つ目のサブテーマにおいて、他のサブテーマでの研究成果を、研究者にサービスするための基盤の研究開発を行い、実際に大学・国内主要研究グループ・学会等に提供しながら、実証的に研究を推進していく。

本研究の実施に先立って行った、第一期新領域融合研究センターおよび次期「新領域融合研究センター」プロジェクト立案のための調査研究において、日本全国の研究者を対象とした研究者向けサイエンス 2.0 基盤サービス「Researchmap」の試行版を公開、運用を始めた。現在までに、300 を超える組織から 1300 人を超える研究者が登録している。4 万を超える論文データが著者本人によって分類・登録されており、本研究が目指す異分野研究資源共有・協働基盤の構築を開始するための下地としては理想的な状態が整っている。また、本サービスの上には、既に 43 の研究コミュニティが構築され、「包括脳支援データベース」「共生社会に向けた人間調和型情報技術の構築（CREST 領域）」など重要な研究コミュニティの共同研究ツールとして実際に活用されている。筑波大学大学院生命環境科学研究科、総合研究大学院大学先導科学研究など、組織としての利用も増加しており、今後、日本の研究サービスの一翼を担っていることが期待されている。

(1) 研究資源に関する情報推薦基盤の構築

研究代表者

〔国立情報学研究所〕　相澤彰子

共同研究者

〔国立情報学研究所〕　内山清子、高須淳宏、宮尾祐介

〔統計数理研究所〕　持橋大地

〔新領域融合研究センター〕　原　忠義

〔広島市立大学〕　難波英嗣

(2) 学術リソースのためのオープン・ソーシャル・セマンティック Web 基盤の構築

研究代表者

〔国立情報学研究所〕　武田英明

共同研究者

〔国立情報学研究所〕　大向一輝、松村冬子
〔国立遺伝学研究所〕　菅原秀明
〔新領域融合研究センター〕　加藤文彦、亀田尭宙、小出誠二
〔東京大学〕　伊藤元己、神保宇嗣
〔人間文化研究機構〕　山田太造
〔東京芸術大学〕　嘉村哲郎
〔ATR-Promotions〕　高橋　徹、上田　洋
〔慶應義塾大学〕　深見嘉明

(3) 多様な知的情報源を結合・融合・再構成する連想情報処理基盤の構築

研究代表者

〔国立情報学研究所〕　高野明彦

共同研究者

〔国立情報学研究所〕　中村佳史、萱島礼香、荒井紀子
〔新領域融合研究センター〕　阿辺川武
〔国際日本文化研究センター〕　丸川雄三

(4) 融合研究を加速するための情報共有クラウドサービスの確立

研究代表者

〔国立情報学研究所〕　新井紀子

共同研究者

〔国立情報学研究所〕　山地一禎、羽田昭裕
〔総合研究大学院大学〕　大田竜也
〔国立極地研究所〕　野木義史、岡田雅樹
〔国立遺伝学研究所〕　菅原秀明
〔新領域融合センター〕　舛川竜治、南　佳孝
〔藤田保健衛生大学〕　宮川　剛
〔電気通信大学〕　Neil Rubens

〔4〕 研究の進捗状況

サブテーマ１

サブテーマ１では、多様な観点に基づき論文を推薦する情報推薦手法の研究およびシステム構築を進めている。平成 24 年度では特に、(1)潜在トピック空間を利用して、分野や言語を横断して論文を推薦する手法の開発に取り組み、論文の抄録に基づき、日本語論文から関連英語論文を推薦するデモシステムを構築した。また、(2)論文の関連研究セクションを言語解析することで、論文中で提案されている「手法」やその「適用対象」を自動抽出する手法を開発した。これらの成果は、研究者による情報探索を支援するための知識獲得基盤技術として今後、実証・活用をはかる予定である。

(1) 国際論文推薦機能の実現

サブテーマ 1 では、研究者に関連論文を提示する情報推薦手法を検討し、その結果をプロトタイプ「オススメ論文検索システム」として実装・評価している（図 1）。「オススメ論文検索システム」は Shibboleth 認証機能を備えていて、Researchmap の ID でログインすると Researchmap ユーザが「公開」に設定した情報をベースにして論文を推薦するなど、将来的には、Researchmap との円滑な連携が可能な設計になっている。

図 1:オススメ論文検索システム推薦結果表示画面

本年度は、昨年度に引き続き「オススメ論文検索システム」を整備し、CiNii 上で公開されている最新の書誌データベースをロードしてコンテンツの充実をはかった。また、昨年度までに実装した「オススメ論文検索システム」の試行的な評価に基づき、今後、公開サービスとしてさらに検討を進める必要がある構成要素の切り出しを行った。現在の実装では、多視点推薦機能として、「速報性」、「類似度」、「人気度」、「異分野」、「入門性」の 5 つのレコメンダを実現していて、ユーザが必要に応じてレコメンダを切り替えられるようになっている。さらに、日本語論文から英語論文を推薦する機能を「国際論文」タグとして実現していて、国際論文の検索結果についても、異なるレコメンダからのランキング結果を表示することが可能である。ここで、国際論文の推薦機能はユーザから高い評価を得ている一方で、現状のシステムの問題点として、(a)和英 2 言語の抄録を持つ論文を経由して言語横断検索を実現しているため、必ずしも言語横断検索の性能が十分でないこと、(b) CiNii 上で提供されている英語コンテンツの数が少ないこと、の 2 点があった。そこで本年度は、国際論文の推薦機能の性能向上に焦点をあてて、(a)平成 23 年度で検討した潜在トピック空間の生成・適用手法を検討するとともに、(b)それに基づく CiNii と海外の論文データベースとの連携の有用性の検証を行った。

一般に言語横断検索を実現する方法として、機械翻訳や対訳辞書などを使って言語間の変換をした後に処理する方法、および複数言語で記述された文書を同一の特徴空間にマップして処理する方法、の 2 通りが考えられる。学術文献は新しい概念や技術を扱うことが多く、新しい用語が使われる頻度も通常の文書より多いため、作成コストの高い対訳辞書等を用いる手法よりも統計的な特性を用いて同一特徴

空間にマップする手法が運用性に優れている。そこで、CiNii 上に登録されている和英 2 言語の論文抄録を利用して、言語混在の潜在トピック空間を生成した。従来の LDA（Latent Dirichlet Allocation）と比較した提案手法の特徴は、1 つの論文から得られる単語を一括して扱うのではなく、メタデータの構造を利用することで、より詳細な潜在構造を抽出している点である。具体的には、論文のメタデータとして与えられる「著者」や「抄録」のレコードの組み合わせに対する生成モデルを構築して、論文間の統計的な類似度を定めている（図 2）。これによって、質的に異なる「抄録中の単語の共起」と「共著者」の情報を統合的に扱うことが可能になる。本年度では、実際に上記の手法に基づき国際論文を推薦するプロトタイプシステムを実装して、トピック数などのパラメタのチューニングを行うとともに、事前計算により実用レベルの応答速度で大規模な推薦サービスが実現できることを実証した（図 3）。

本研究で検討した情報推薦の枠組みは、言語横断検索だけではなく、電子図書館サービスの横断的な連携に一般に適用可能である。電子図書館サービスどうしの連携においては、オープンな形でメタデータをやりとりすることが基本となるが、例えばアクセス数、引用、本文コンテンツ、実験データなど、各電子図書館が付加価値として論文に追加している情報については、必ずしもオープンな形でのやりとりが可能ではない。潜在トピック空間は、個々の電子図書館が独自に構築している論文空間どうしをつなぐ役割を果たすもので、共通して登録されている論文や専門用語など共有する少数のリソースから潜在トピック空間を構築することで、電子図書館どうしの独立性を保ちつつ相互にトラフィック誘導が行えるという利点がある。また、国内の研究者を海外ジャーナルに積極的に誘導する仕組みをアピールすることで、海外の最新のコンテンツが入手しやすくなることも期待されるため、今後も国際論文レコメンダの性能向上を目指して検討を進める予定である。

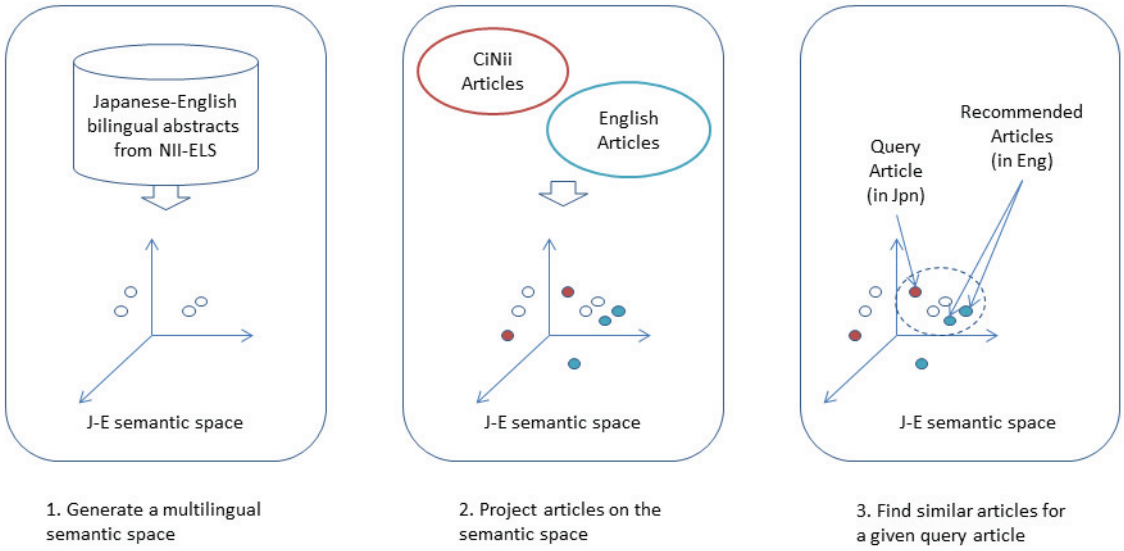

図 2：潜在トピック空間を用いた国際論文推薦機能の実現

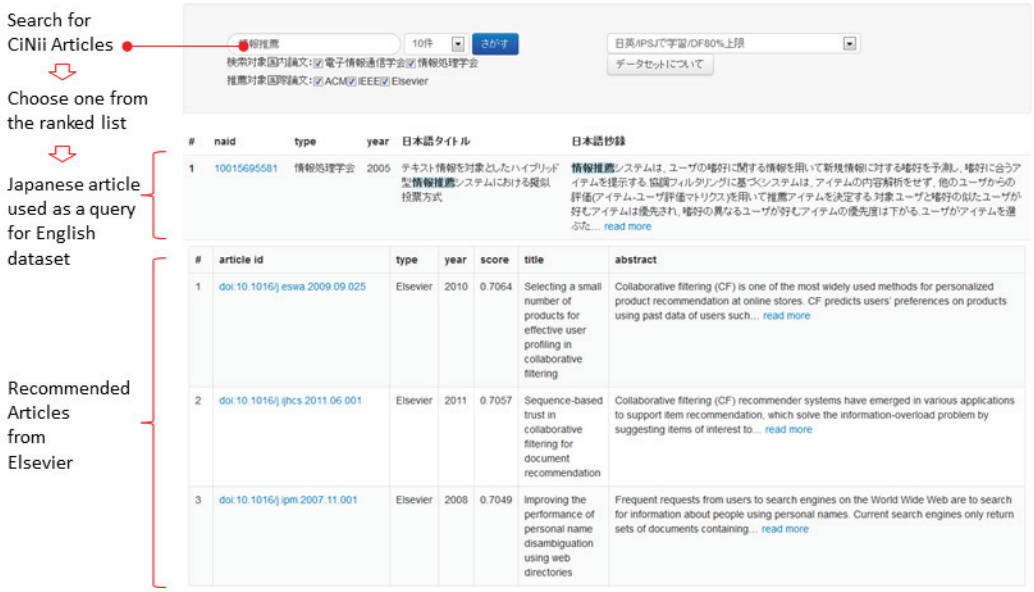

図 3：国際論文レコメンダ

(2)　論文中の「手法」や「適用対象」の自動抽出

論文内で記述されている引用文献情報、研究で用いられている手法などの情報を集約し、リンクで関連付けた論文知識ネットワーク構造を提示することにより、研究者の研究活動支援を目指す。具体的には、論文内で記述されているアルゴリズムや手法などの専門用語（e.g. “PHITS”, “PLSA”）、名詞句（e.g. “テキストのリンク情報”）、動詞句（e.g. “語の共起行列を生成する”）、節など、論文知識を表現するに当たって有用な文以下の単位を纏めて「概念（Concept）」と呼称して抽出し、これらと引用されている「論文」を結び付けることで、「論文」と「概念」の 2 種類のノードによるネットワークを構築することを目指す（図 4）。このネットワーク構築により、「同じテーマを扱っている論文を収集し、論文内から抽出される論文―概念間、概念―概念間の関係をまとめることで、その分野全体の先行研究のネットワークが概観できる」、「同じ論文を引用している論文にある関係を統合することで、先行研究の様々な取り組みについて理解を深められる」さらに、「分野を概観することで、自身の研究を周辺分野および関連研究の中に的確に位置づけやすくなる」などの研究活動支援効果が期待できる。

このような論文―概念間、概念―概念間の関係には、以下の 6 種類があると考えられる。

・Method—Purpose: 手法と目的、応用先
・Topic—Paper: 概念とそれが扱われている論文
・Method—Purpose: Method-Purpose の否定
・Citing—Cited: デジタルライブラリから取得することのできる論文同士の引用-被引用関係
・SameAs: 同義語
・Super—Sub: 上位下位もしくは全体部分関係

本研究では、このうち Topic—Paper と Method—Purpose の関係を英語論文から抽出することを試みた。関連研究（related work）の章を対象とし、各文に構文解析ツールを適用して得られた構文構造に対して、以下のような人手で作った名詞句（NP）、動詞句（VP）、および “[2]” のような引用文献符号（CITE）の間の構文パターン 15 種類を用いて、15 論文から関係の抽出を行った。

・ “({NP1}) {be} based on ({NP2})” → NP1 = “Purpose”, NP2 = “Method”
・ “({VP}) using ({NP})” → VP = “Purpose”, NP = “Method”
・ “({NP}) ({CITE})” → NP = “Topic”, CITE = “Paper”

評価に人手のアノテーションデータが必要な実験のため、抽出対象とした論文数は少ないものの、構文パターンにより高い精度・再現度で関係抽出が行えることが確認された。一方で、分詞の扱いや、副詞の挿入、文・節の末尾にある引用文献符号の扱いなど、現在の構文パターンでは扱いきれない現象や、構文解析ツールが誤った解析結果を出力する場合など、抽出に失敗するケースも時折見られるが、それらに関しては、後処理や構文パターンの追加によって改善が期待されるため、今後、より大規模な論文データを用いた関係抽出に取り組む中で取り扱って行きたい。また、抽出した知識の統合に取り組むことで、論文・概念間の関係に基づいて分野全体としての論文知識ネットワークを表現したいと考えている。

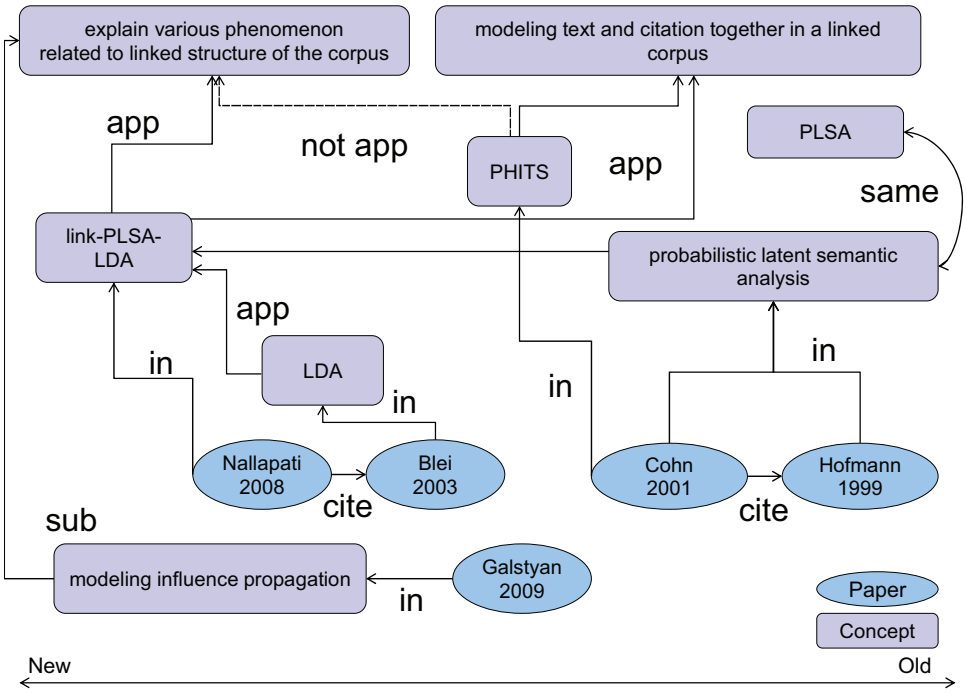

図 4：概念―論文ネットワーク

上記のネットワークによるマクロな研究分野の全体像を捉えた上で、次の段階として、各論文の中でより具体的にどのような記述（目的・手法・貢献）がなされているかという、より深い情報探索を支援すべく、アブストラクトの各文の記述から、その記述と対応した詳細記述のある論文本文パラグラフを自動的に獲得する（図 5）2 手法を提案し、その可能性を検証した。1 つは、アブストラクトの文と本文パラグラフの最長共通部分列（Longest Common Subsequence, LCS）に注目し、語の一致度と各単語の重要度からもっとも近いと思われるパラグラフを決定する手法で、もう 1 つは Latent Dirichlet Allocation（LDA）というトピックモデルを用いて、トピック分布の類似度から最も近いパラグラフを決定する手法である。前者は、アブストラクトと本文で全く同じ言い回しが観察されるという側面に注目した手法で、後者は、全く同じ言い回しではないものの、近い表現を用いたパラグラフを発見することを目指したもので、一般に文書単位で適用される LDA を、より狭いパラグラフ単位でのトピックの近さの指標に用いた形となっている。

実験では、30 論文に対して、アブストラクトの各文と本文パラグラフとの対応獲得を試みた結果、絵前者のモデルが後者のモデルよりも高い精度で対応を獲得できた。後者のモデルに関しては、人手でアノテーションを行った学習データが圧倒的に不足していたこともあり、高い精度は出せなかったものの、前者のモデルで見つけられなかったパラグラフとの対応が獲得できるケースも見られ、実験規模の拡大による性能改善が期待される。

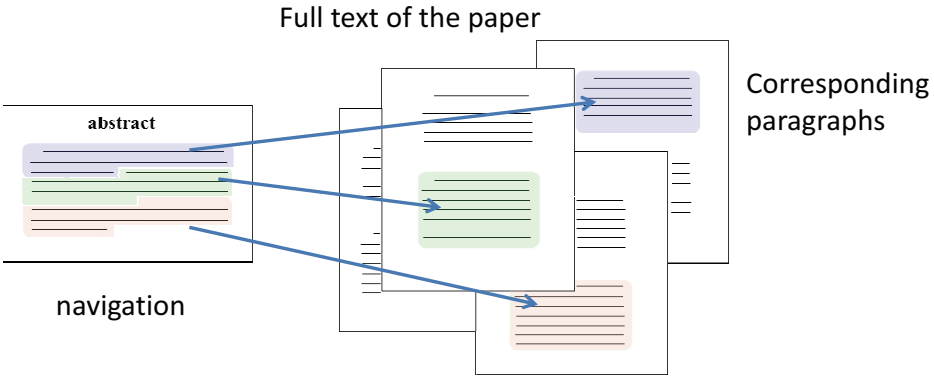

図 5：論文アブストラクトの各文から本文パラグラフへの対応づけ

サブテーマ 2

1. はじめに

学術リソースのためのオープン・ソーシャル・セマンティック Web 基盤の構築プロジェクトにおいては、多様かつ分散的なデータを柔軟に連携させる仕組みについて Linked Open Data の方法論に基づき研究を進めている。

Linked Open Data は、RDF などの言語を用いて記述されるシンプルで柔軟性がある仕組みで、多様なデータを記述することができる。そのため、欧米や米国では、既に新しい情報公開・共有の仕組みとして Linked Open Data が認知されつつあり、情報流通の仕組みとして普及しつつある。また、我が国においてもさまざまな研究や活動が行われている。

本研究では、生物学の中でも生物多様性の分野に焦点をあてている。この分野は、現在、生物多様性の損失や保全など、地球環境問題の 1 つとして社会問題にもなっている。これらの問題を解決するには、対象生物のみではなく、地球規模の観測から人間活動まで様々な情報を横断的にデジタル・アーカイブとして利用できる基盤が必要である。しかし、生物の種名や分布、各種の特徴や保全状況といった生物的な情報でさえ、現状では形式や公開場所が分散しており関連が弱い。そこで、本研究では、Linked Open Data の技術を用いて、分散的に公開されている生物多様性の情報を統合的に利用できるようにすることを考えた。

昨年度において、生物種の情報についてそのモデル化と LOD 作成を行った。また一昨年度には博物館・美術館のデータについて LOD 化を行ってきた。そこで、今年度は、この二つを連携させ、博物館などの収蔵されている標本データについて両方の LOD につながる形での LOD 化を行う。

2. S-Net について

生物には、分子レベルから生態系レベルまで多層のレイヤーが存在し、生物多様性もこうした多層レイヤーから構成されている。

中でもその中核をなす種の多様性は、主に個体や種の名称・特徴といった情報が扱われ、大きく分けて (1)生物名の目録の情報（種名情報）、(2)標本や観察記録などの情報（分布情報）、(3)それぞれの生物種の特徴を示す情報（種情報）からなる。このようなデータを情報技術により保存・解析・活用

することを目的とした横断的分野は、生物多様性情報学（biodiversity informatics）とよばれる。

このような生物多様性情報は、生物分類学の研究成果として、18 世紀より紙媒体に蓄積されてきたが、情報技術が発達した現在では、膨大な情報を扱うデータベースに重要な情報ストレージとして蓄積されている。その例としては、グローバルなものとして地球規模生物多様性情報機構（The Global Biodiversity Information Facility：GBIF，種名・分布情報）、Encyclopedia of Life（EoL，種情報）、Catalogue of Life（CoL，種名情報）、Barcode of Life Data Systems（BOLD，DNA・標本情報）などが挙げられる。一方、国内では国立科学博物館が運営するサイエンスミュージアムネット（S-Net，標本情報，GBIF と連携）がある。

S-Net は、全国の科学系博物館のポータルサイトとして全国の科学系博物館の情報を横断検索できる Web サイトである。S-Net では、各博物館の Web ページ上にある情報を検索できる「Web 情報検索」と、各博物館が保有する標本情報と採集に関する情報を検索できる「自然史標本情報検索」の 2 種類の検索を行うことができる。特に、「自然史標本情報検索」では、全国の 55 館の協力博物館から収集した情報を一意の形式で提供しており、人文科学の観点からも、利用者に有益なサービスとなっている。

しかし、個別の情報を見てみると、多数の博物館から収集されているため、分類群に関するデータが記載されていないなど情報の偏りや表記揺れがある。また、その情報は、人間が読むことを前提に作られており、コンピュータが利用しやすい形式になっていない。

そこで、本研究では、S-Net で提供されている標本情報を対象に、Linked Open Data 化することによって、それらの問題点を改善することを考えた。

3. 標本情報の Linked Open Data 化

S-Net で扱われている標本情報は、学名、一般名（和名）、分類群に関する項目、種の命名者、採集地、最終日、採集者番号、標本の性別、グローバルユニーク番号、データ種別、タイプ標本、所蔵博物館、備考というデータで構成されている。また、S-Net が提供しているサービスは、検索サービスのみであるため、クローリングによってデータを取得する。

一方、これまで、筆者らが LODAC プロジェクトで取り組んできた生物多様性情報は、学名、著者、出版年、和名、和名の別名、分類群に関する項目を対象としている。

前述したように、S-Net の情報は、分類群に関するデータが記載されていないなど情報の偏りがあるため、そのデータについては、LODAC プロジェクトで構築したデータを利用する。S-Net に記載されている標本 1 件につき 1 つの URI を生成し、一般名（和名）、採集地、採集日、所蔵博物館のデータに対して Linked Open Data 化する。

さらに、LODAC プロジェクトでは、すでに博物館情報を機関 URI として保有しており、標本が所蔵されている博物館を既存のデータにリンクすることも行う。

具体的には、図 1 に示すデータ構造に基づいて、Linked Open Data 化を行う。まず、各標本情報に LODAC プロジェクトで定義している固有の ID を割り当て、URI を生成する。URI には、それが標本情報であることを判別できるように、rdf:type の定義を行う。そして、生成した URI から種情報へのリンク、S-Net の情報が記載された Web ページへのリンク、LODAC プロジェクトで定義した機関 URI へのリンクを生成し、一般名（和名）、採集地、採集日、所蔵博物館名については、リテラル（文字列）のリンクを生成する。

4. Linked Open Data 化の問題と対処

S-Net のデータは、全国の協力博物館から情報を収集しているという性質のため、前述した (1)分類群に関する情報の有無の他に、(2)学名の表記揺れや (3)所蔵博物館名の記述方法の違いがあった。これらの問題について、どのような事例があるのかを確認した上で、Linked Open Data 化する場合

の対処法を考えた。

(1) 分類群に関する情報の有無

まず、分類群に関する情報の有無については、界名から種小名まで全て記載されているデータや属名・種小名のみといったデータなど、様々な場合があった。一方、LODAC Species では、これまでに種名情報と分類群に関する情報を構築してきた。

そこで、S-Net の標本情報から LODAC Species の種名 URI へのリンクを適切に記述すれば、S-Net のデータを Linked Open Data 化したあと、標本情報から分類に関する情報まで参照することができると想定し、S-Net で記述されている分類群に関するデータについては今回の Linked Open Data 化の対象から除外した。

(2) 学名の表記揺れ

次に、学名の表記揺れについて対応を考えた。S-Net で提供されているデータについて、Papilio xuthus（和名：アゲハチョウ）の例を挙げると、学名である「Papilio xuthus」と記述されている場合の他に、「Papilio xuthus Linnaeus」や「Papilio xuthus Linnaeus, 1767」、「Papilio xuthus xuthus LINNAEUS, 1767」、「Papilio xuthus B1292175」といった形式で記述されていた。学名は、属名（Papilio）と種名（xuthus）を組み合わせて表記する 2 名法が国際命名規約に基づく生物種の命名法として一般的であるが、「Papilio xuthus Linnaeus」や「Papilio xuthus Linnaeus, 1767」などは、学名の後ろに命名者や年号が付加されたものである。また、「Papilio xuthus B1292175」という表記は、博物館で管理されている識別番号を学名の後ろに付加したものである。この他にも、命名者や年号が括弧で括られている表記などもあり、学名だけ見ても多様な表記揺れが存在する。

この問題に対して、多様に表記されている学名全てに URI を割り当て、owl:sameAs で種名 URI にリンクする方法を採用した。この方法を採ることによって、オリジナルのデータを改変する必要無く、既存のデータとリンクする事ができ、種名とリンクしている分類群などの各種データを参照することができるようになる。

(3) 所蔵博物館名の記述方法の違い

S-Net の所蔵博物館名を確認すると、LODAC Museum でこれまで扱ってきた美術館名や博物館名と、標本情報が所蔵されている博物館名の記述方法が異なることがわかった。例えば、LODAC プロジェクトでの博物館名が「北九州市立自然史・歴史博物館（いのちのたび博物館）」となっているのに対し、S-Net のデータでは、「北九州市立自然史博物館」と表記されていた。また、国立科学博物館を指すと思われる名称が LODAC Museum では、「国立科学博物館」、「独立行政法人国立科学博物館附属自然教育園」、「国立科学博物館産業技術史資料情報センター」、「国立科学博物館分館」と複数存在し、S-Net のデータにも「国立科学博物館 植物研究部」、「国立科学博物館（動物）」、「国立科学博物館（動物・人類）」、「国立科学博物館（植物）」と複数存在することがわかった。このような記述方法は、博物館が、組織や施設、コレクションと言った目的ごとにデータセットが異なることから発生したと考えられる。

Linked Open Data 化において、リテラルでそのままデータを登録するには問題は無いが、これらのデータをリンクして活用することを考えると、非常に利用しにくいデータとなる。そこで、LODAC Museum で管理している美術館名や博物館名を含む機関 URI と比較し、該当する博物館の有無を確認した。該当する博物館が無い場合は、所蔵博物館について機関 URI として新規追加し、機関 URI のデータとリンクする事を考えた。

LODAC Museum と LODAC Species では、それぞれで SPARQL endpoint を公開している。そこで、この 2 つの SPARQL endpoint を活用してリンクを生成する事を考えた。この処理には、Silk というツールを利用した。

Silk は、2 つの異なるデータソースのデータ項目間のリンクを生成するツールである。SPARQL endpoint を利用でき、リンクするプロパティはもちろん、2 つのデータソースを比較・マッチングするときに、様々な条件を設定することができる。

本研究では、LODAC Museum と LODAC Species の SPARQL endpoint を用いて、LODAC Museum の約 20 万件と LODAC Species に登録した約 120 万件のデータを対象にリンクを生成した。合計で約 2400 億回のマッチング処理が行われ、処理時間は約 11 時間であった。

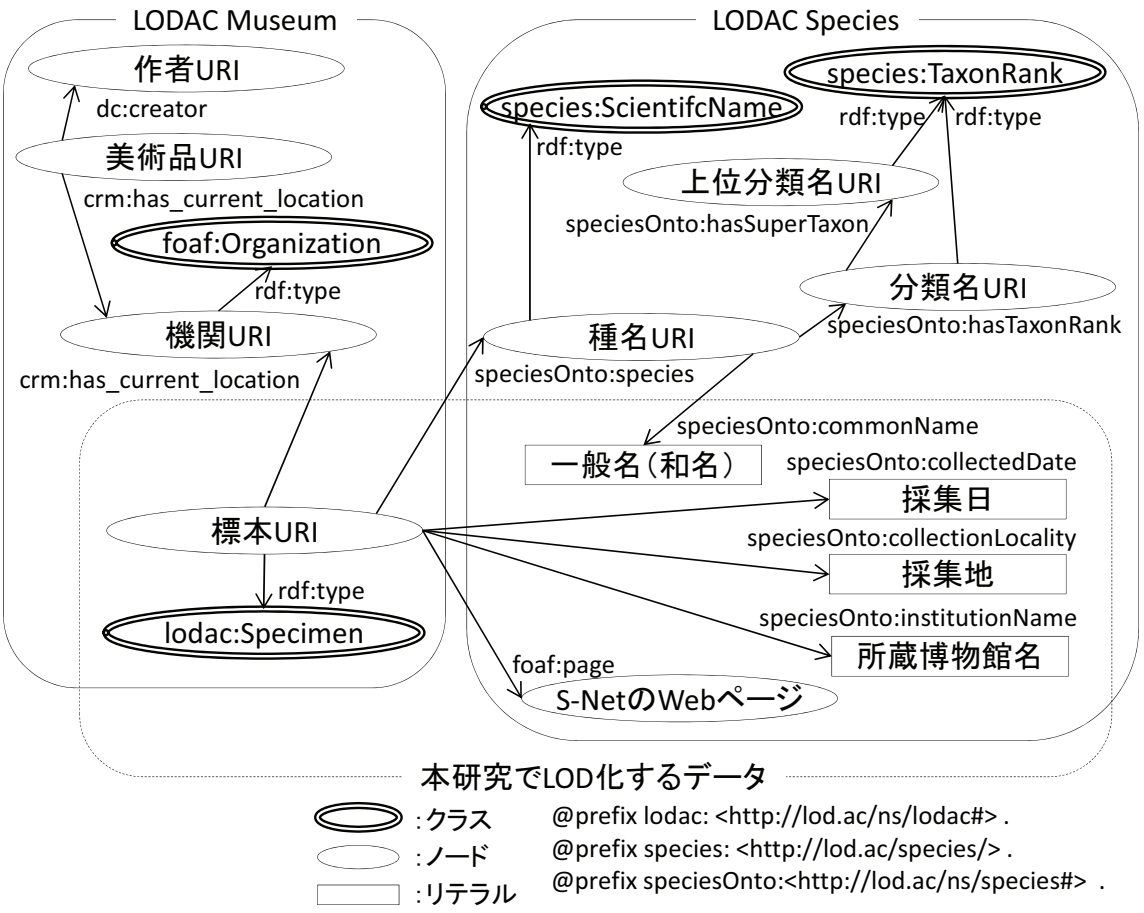

図 1：データ構造

5. Linked Open Data 化の結果

上記の方法で、S-Net のデータを Linked Open Data 化し、LODAC プロジェクトの RDF Store に登録した。トリプル数は、9,754,474 となった。

LODAC プロジェクトでは、HTML によるインターフェースを備えており、SPARQL Endpoint も公開している。そのため、個別の標本情報を閲覧することはもちろん、標本情報に関連する情報についてリンクをたどって閲覧したり、ある種に関する標本情報を一覧として取得したりすることができる。

図 2 に、アゲハチョウに関する結果例を示す。左側が LODAC Museum に登録したアゲハチョウの標本の内容である。ページ内の左側の列がプロパティ、右側の列がオブジェクトを示している。1 行目は、rdf:type で、このデータが標本データであることを示している。2 行目は、S-Net から取得した学名をラベルとして表示している。3 行目は、図 2 の右上に示す LODAC Species のアゲハチョウの種名 URI へのリンクを示している。種名 URI には、分類群や関連 Web ページへのリンク、画像な

どがリンクされており、それらのデータを閲覧することができる。4 行目と 5 行目は、採集日、6 行目は採集地を示している。7 行目は、S-Net がデータソースであることを示している。8 行目と右下は、データソースとなった S-Net の Web ページを示している。9 行目と 10 行目は、LODAC Museum での管理上の ID を示している。11 行目は、図 2 の右下に示す S-Net へのリンクを示しており、データを比較すると、同一のものであることがわかる。12 行目は、標本の所蔵博物館を示している。

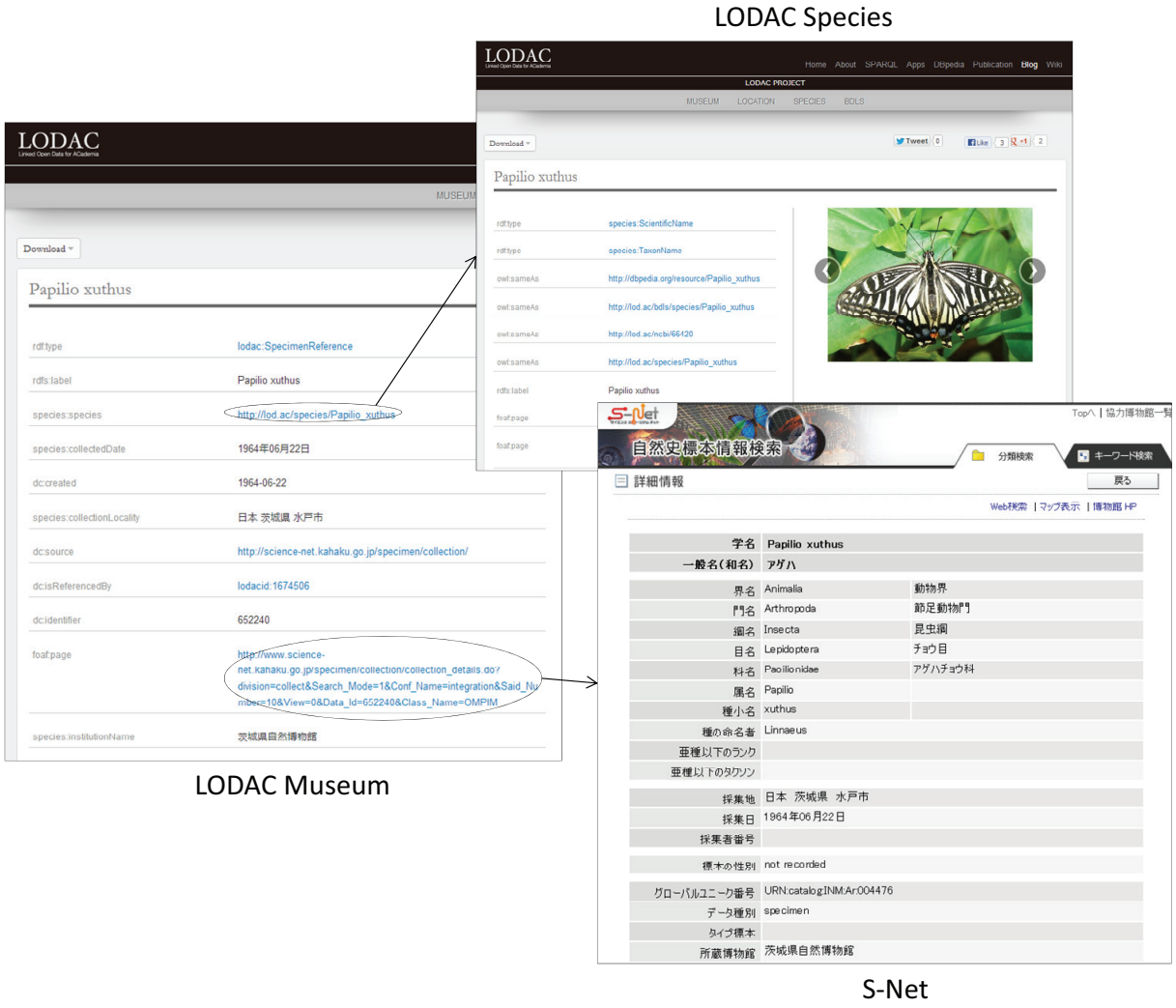

図 2:結果表示例

6. おわりに

本研究では、S-Net で提供されている標本情報を対象に、Linked Open Data 化を行った。その結果、LODAC プロジェクトでこれまで構築してきたデータとリンクすることができ、情報の偏りや表記揺れを補完する働きを実現できた。

分類群に関する情報については、LODAC Species とリンクすることによって、S-Net のデータに欠けていた情報を補完することができた。これにより、SPARQL endpoint を利用して、同属の標本を抽出するといった検索も可能になった。

学名の表記揺れについては、それぞれの学名に URI を付与することによって、関連する学名で異なる表記のものを一覧することができるようになった。

所蔵博物館名の記述方法の違いについては、機関 URI のデータとリンクする事によって、異なる目的のデータセットを繋ぐことができた。これにより、「美術品」と「生物標本」といった異質なデータに関連が生まれ、人文系と自然科学系の博物館の所蔵品をシームレスに扱うための基盤となり得ると考えられる。特に、人文科学における考古学資料と自然科学系の標本資料は、どちらもその資料を採取した場所や日時、所蔵機関などが連携できると考えられ、そこからつながる様々なデータのハブになると考えられる。

このように、生物標本情報を Linked Open Data 化し、これまでに構築してきた情報とリンクすることによって、コンピュータが利用しやすい形式になった。そのため、要求に応じた柔軟な検索が可能になり、未知なデータとつながる基盤としても今後、大きな可能性があると考える。また、本研究の成果は、デジタル・アーカイブとして標本情報の利用価値の向上、そして、相互運用性の向上にもつながると考える。

サブテーマ 3

1. はじめに

本年度は、連想検索技術を活用したコンテンツ表示システムの開発、および連想検索システムに時間や空間の概念を取り入れた実運用サービスの開発をおこなった。

2. コンテンツ表示システムの機能追加

書籍や報告書などの主にテキスト情報から構成されるコンテンツについて、テキストの理解を促進するために文中のキーワードに対する百科事典や辞書の項目をサイドノート（脚注部）に表示するシステムを開発している。本年度は、脚注部に表示する情報として、百科事典や辞書のようなあらかじめ定義された見出し語集合を持つリソースだけでなく、外部のサイトが提供するサービスとの連携機能を追加した。

(1) 連想検索を用いた情報提示

従来の脚注表示機能は、本文中から抽出したキーワードを用いて、システム内部で管理するリソースを表示していたが、表示している本文全部を用いて外部のサイトが提供する連想検索を利用する仕組みを開発した。プロトタイプシステムでは外部サイトの情報として、想 IMAGINE の提供する「新書マップ」「文化遺産オンライン」の情報を連想検索で検索し、本文と関連の深いコンテンツを表示する（図 1 の左右の一番外側の列は新書マップを用いた連想検索結果）。

(2) 動画配信サイトを利用した動画へのリンク機能

脚注表示機能では、キーワードに関する説明文の他に写真を表示することで、文字だけから構成される本文に対し、キーワードの具体的なイメージを提示することができる。本年度はさらに、外部の動画配信サイトが提供する動画検索 API を利用して、キーワードに関連する動画をリアルタイムに検索し、動画を紹介する写真とその動画へのリンクを表示する機能を追加した。動画検索 API を利用したキーワード検索では、多義語を持つキーワード（例えば「オアシス」は場所、バンド名、ワープロ名など）の場合、本文と関連のない動画が検索されることがあるため、本文中で用いられている意味を推定し、動画検索時に推定した意味と関連するカテゴリーに限定する指定をしている。

図 1 では動物のイラストがすでに本文に掲載されているが、脚注部の一番内側の動画へのリンクを辿ることで実際に動いている動物を見ることができる。このように図鑑や教科書、資料集など子供向けのコンテンツに適した機能といえる。

図 1:映像検索を取り入れたコンテンツ表示システムの例

3. 時空間を考慮した連想検索システムの開発

従来の連想検索方式では、テキストから抽出される単語から構成される語空間を検索空間ととらえていたが、ここに時間、空間の概念を取り入れた。指定した時間や空間の近傍を検索する既存のシステムでは、検索クエリーとして 1 点のみしか入力できないという制約があるが、連想検索では複数のコンテンツを検索クエリーとして入力できることが特徴の 1 つであるため、時空間の検索においてもそれを可能とする技術を開発した。具体的には、各コンテンツの持つ時空間の値を平均とする正規分布を作成し、複数コンテンツが入力クエリーの場合は時空間を確率分布の和として表現し、この値を離散化したベクトルを連想検索に用いた。以降では、時間、空間を考慮した連想検索を使用し、すでに運用されているシステムについて報告する。

(1) 時間を考慮した連想検索

NHK 放送文化研究所では、NHK の発行する年鑑、年報、月報、放送史年表といった各種調査資料の作成を行っており、NHK 開局以来、大量の資料がアーカイブとして保存されている。現在、これらをスキャンし OCR を施す電子化を進めている最中であるが、ある資料と関連する資料を探すといった、単純な文字列検索だけでは実現できない検索をしたいという要望があり、我々は連想検索を使った検索システムの開発の協力や、先に述べたコンテンツ表示システムをアーカイブの閲覧機能として提供している。

その中で 20 世紀放送史年表という NHK の設立当初からの放送に関する出来事を年表として列挙したコンテンツがあり、出来事とそれに関するアーカイブ資料を検索する仕組みを開発した。年表項目の多くは短い 1 文で記述されていることが多く、構成する単語数が少ない場合、連想検索が有効に機能しないことがある。そこで年表項目に付与されている時間情報を利用し、項目中の単語と時間を考慮した連想検索機能を使用することで、内容が関連し、かつ時間的にも発行年の近いアーカイブ資料を検索できるようになった（図 2)。

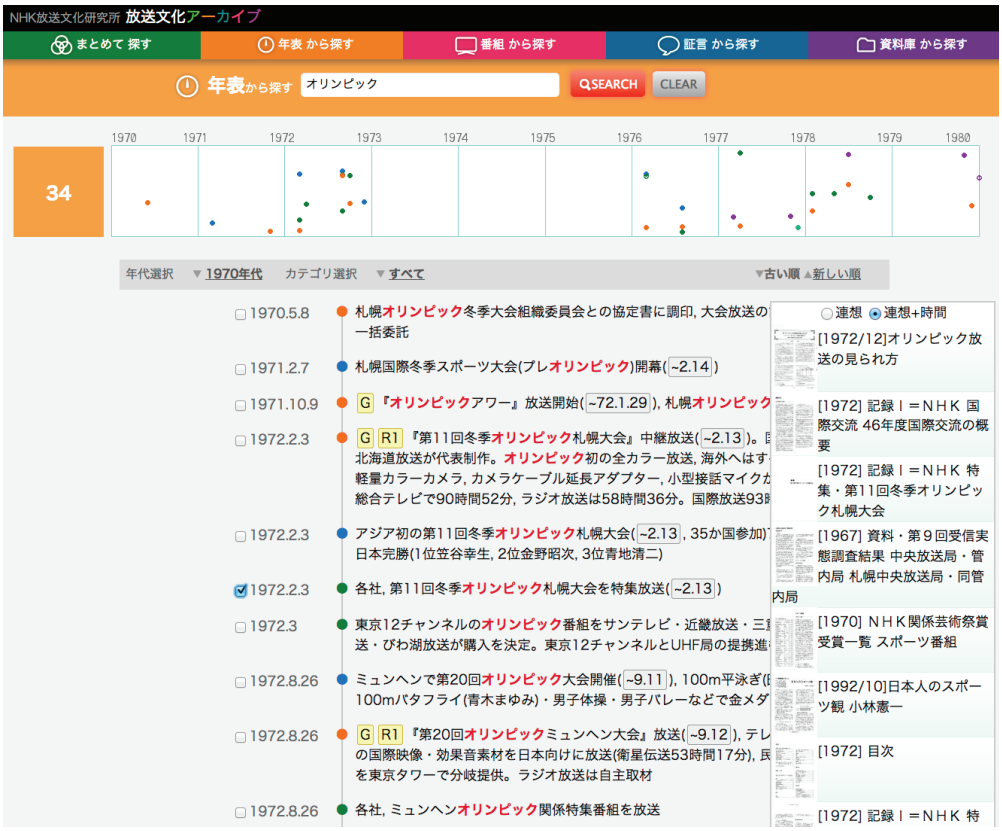

図 2:時間情報を考慮した連想検索の例

(2) 空間を考慮した連想検索

2013 年 4 月に御茶ノ水ソラシティ内に街歩きの起点としての案内所の機能を持つ「お茶ナビゲート」がオープンした。この中で、お茶の水界隈の自分のお気に入りのスポットを探してオリジナル散歩地図を作成できるサービスが稼働しており、このシステム内で空間を考慮した連想検索機能を使用している。ユーザはお気に入りのスポットを選択し、システムはユーザの興味に近く、かつ徒歩での移動を想定しているため場所的にも近いスポットをおすすめスポットとして提示する（図 3)。時間の場合は 1 次元の確率分布であったが、空間の場合は 2 次元の確率分布を生成し、選択スポットの確率分布の和を離散化したベクトルと、検索対象であるスポットが持つ場所ベクトルとの類似度を計算する処理をおこなっている。

図 3:空間を考慮した連想検索の例(オレンジ色が選択したスポット、青色がおすすめスポット)

今後は、文化遺産や東日本震災アーカイブなど、あらかじめ時間や場所情報がメタデータとして付与された大規模なデータについて、時空間を考慮した連想情報技術を適用し、機能の有効性を検証していく予定である。また時空間情報は、テキスト文中で言及されていることも多く、自動的に抽出した時空間データを考慮した時空間連想検索と、従来の単語空間だけの連想検索との検索精度の比較などもおこないたい。

サブテーマ４

1. はじめに

サブテーマ１から３においては、テキストを中心とした大規模データの格納・検索・共有・分析・可視化に関する要素技術を研究しているが、サブテーマ４では、Researchmap という 22 万人の日本の研究者情報という具体的な大規模データを基盤として、それらの要素技術を検証し、その具体的な課題をフィードバックする研究サイクルの確立を目指している。同時に、本テーマで構築した研究情報のデータベースと共同研究基盤がそれ自体として、日本の共同研究、特に学際的な融合領域の研究

が促進することを目指す。

コンピュータ群に研究情報をより高精度でより効率よく処理させるためには、さまざまな観点から機械が可読であるような情報アーキテクチャを十分に検討する必要がある。たとえば、印刷された論文をスキャンして画像として保存された文書と、タイトルや著者氏名、キーワード等にアノテーションを施し、実験データにはローデータの所在などがリンクとして埋め込まれた形式の文書では、人間の見た目には同じように映る「デジタル化された論文」であっても、その機械可読性には雲泥の差がある。特に、深い検索、深い分析を可能にするには、粒度がそろった正しい情報が大量に蓄積されていることと、機械による処理を可能にするデータベースの設計と厳密な情報の入力が求められる。しかし、その一方で、機械可読性を求めるあまり厳密な情報の入力をユーザに求めれば、入力コストが効用を上回る。生命科学の文献情報を収集したオンラインデータベース Medline では、正確な検索を保証するためにデータ整備に毎年膨大な人件費を支出している。一方、完全な自動化を目指した情報科学分野の文献情報システム CiteSeer では、検索精度が極めて低い。このことからも精度とコスト削減の両立がいかに困難な課題かがわかる。

どのようなプラットフォーム（制度）を導入すれば、機械可読性やユーザの効用の向上と、コストの縮減を同時に実現しうるだろうか。本サブテーマでは、研究資源が発生時点からデジタルであるようなボーンデジタル時代の学術研究情報のエコシステム（循環型情報活用基盤）を今後いかに確立すべきかについて検討を行い、Researchmap という基盤ソフトウェアとして実装していく。

2. 異分野共同研究基盤システム Researchmap の機能と構成

Researchmap の機能を要約すると以下のようになる。

1　PubMed，Amazon，CiNii, KAKEN, J-Global など標準的な規約（RSS, ATOM 等）に基づきデータをオープンに公開している学術データプロバイダーから、研究者の研究業績・競争的資金の獲得状況などをフィードできる。

2　研究者リゾルバー[1]を用いて研究者の名寄せを行い、高い精度で研究業績等のフィードが行える。

3　1 および 2 の機能を用いて、研究者が半自動的に本文所在情報の URL を埋め込んだ研究業績リストおよび Curriculum Vitae（CV）を備えたウェブページを作成し、このページに対して不変の URL を付して、研究者にホームページサービスとして提供する。

4　登録研究者が「自分の業績」として認めた項目情報を CiNii や J-GLOBAL 等のデータプロバイダーにフィードバックすることにより､機械だけでは困難であった研究者情報の名寄せを補完する。

5　登録研究者は CV ページ以外にも、研究ブログを公開したり、研究・教育資料のリポジトリ機能を利用したりすることができる。

6　CV に登録したデータはテキストまたは CSV 形式でダウンロードでき、各種の申請や報告書（大学評価・年報・競争的資金の応募書類・報告書)、調査等に流用できる。

7　Researchmap の CV データは ATOM1.0 に準拠した形式で大学等の研究機関に提供する。

8　登録研究者は学術･研究イベントを登録し、広報することができる。登録されたイベントは、RSS 形式で配信される他、twitter 上でロボットが情報を拡散する。

9　研究プロジェクトを推進するためのバーチャルラボ（コミュニティ）機能を提供する。コミュニティ機能には以下のツールが予め搭載され、利用することができる。

9.1　メーリングリストと連動した掲示板機能

[1]NII が提供している研究者の情報を集約してアクセスを可能にするためのサービス。

9.2　データを共有するためのキャビネット機能および汎用データベース機能
9.3　ToDo を管理するための ToDo モジュール
9.4　予定を調整・管理するためのスケジューラー機能およびカレンダー機能
9.5　オンライン会議のためのチャット機能
9.6　プロジェクト内でアンケートを取るためのアンケート機能および投票機能
9.7　リンクを共有するためのリンクリスト
9.8　写真・画像を整理するためのフォトアルバム機能
9.9　登録研究者間で個人的なメッセージをやりとりするためのプライベートメッセージ機能

10　Researchmap 上で起きている情報をパーソナライズした上で整理・分類して可視化する「新着情報」機能を利用できる。

11　以上の機能のうち CV 以外の機能を携帯電話で閲覧・編集できる。

12　登録研究者を名前・所属・研究分野・研究キーワード・所属学会・地域などから多角的に検索できる。

13　登録研究者の研究内容上の距離を定義し、関連研究者（おとなりの研究者）を計算し、可視化する。

以上の機能を研究者に提供することで、自然科学から人文科学にわたる異分野の「知」と「人」の共有・連携を行い、情報や研究人材の効果的な活用や研究協力・共同研究の促進を行うボーンデジタル時代の学術研究情報のエコシステム（循環型情報活用基盤）を構築する（図 1）。

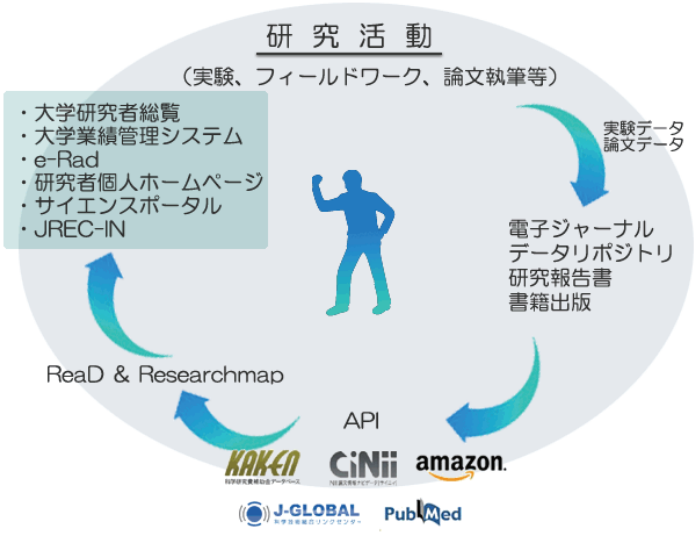

図 1：学術研究情報のエコシステムのイメージ図

3.　平成 24 年度の成果

Researchmap の機能のうち 7 である「CV データの ATOM1.0 に準拠した形式での API」を設計し、公開した。これにより、情報・システム研究機構の女性研究者総覧（羽ばたけ日本の女性研究者 http://women.rois.ac.jp/）の全自動構築を可能にした。他に、北海道大学・高専機構など 30 機関に API を提供しており、Researchmap の API を活用した研究者総覧が各地で構築されている。

また、本プロジェクトを始めとして、各種学術・研究サービスとの ID 連携を図るため、Shibboleth

による認証の仕組みを検討しこれを構築した。Shibboleth を用いた Researchmap との ID 連携は、e-Rad を始めとして 13 の機関に広がっている。

多種データのリンケージ・可視化に関する研究のデータを構築するため、研究業績に加えて新たに特許情報のフィード機能を検討・実装した。これにより、従来無関係に存在していた研究情報と特許情報を Researchmap 研究者 ID を介して紐付けることが可能となった。

図 2：Researchmap の API を活用して構築された女性研究者総覧

次期 Researchmap のコアシステムとして、オープンソースのコンテンツマネージメントシステム NetCommons3.0 の開発を開始した。NetCommons は Researchmap の API を活用して各大学の研究者総覧を作成する上でも基盤となるシステムでもあり、高専機構・北海道大学・筑波大学をはじめとして 3500 の教育機関で活用されている他、サブテーマ 1 において鍵となるクラウド上の機関リポジトリシステム weko の基盤システムでもある。NetCommons の柔軟性とセキュリティを高め、スマートシステム時代のインタフェイスに対応することが本研究で開発されている各種技術が継続して研究者に利用される上で不可欠である。特に、スマートシステムに対応することで、フィールドワーク研究との接合が期待される。

〔5〕 研究成果物

＜論文発表＞

〔学術論文〕

1. Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa: “Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet”, The 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012) pp.686-695 (2012.04)

2. Hidetsugu Nanba, Toshiyuki Takezawa, Kiyoko Uchiyama and Akiko Aizawa: “Automatic Translation of Scholarly Terms into Patent Terms Using Synonyms Extraction Techniques”, The 8th International Conference on Language Resources and Evaluation (LREC 2012) pp.3447-3451 (2012.05)
3. Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa: “Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure”, The 8th International Conference on Language Resources and Evaluation (LREC 2012) (2012.05)
4. Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Sophia Ananiadou and Akiko Aizawa: “Normalisation with the Brat Rapid Annotation Tool”, The 5th International Symposium on Semantic Mining in Biomedicine (SMBM 2012) (2012.09)
5. Pascual Martinez-Gomez, Tadayoshi Hara, Chen Chen, Kyohei Tomita, Yoshinobu Kano and Akiko Aizawa: “Synthesizing Image Representations of Linguistic Features over a Text for Predicting Areas of Attention in Reading”, The 12th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2012) pp.312-323 (2012.09)
6. Giovanni Yoko Kristianto, Goran Topic, Minh-Quoc Nghiem and Akiko Aizawa: “Annotating Scientific Papers for Mathematical Formulae Search”, Proceedings of the 5th workshop on Exploiting semantic annotations in information retrieval (ESAIR 2012) of The 21st ACM International Conference on Information and Knowledge Management (CIKM 2012) pp.17-18 (2012.11)
7. Tadayoshi Hara, Daichi Mochihashi, Yoshinobu Kano and Akiko Aizawa:“Predicting Word Fixations in Text with a CRF Model for Capturing General Reading Strategies among Readers”, the Workshop on Eye-Tracking and Natural Language Processing (ETNLP) of 24th International Conference on Computational Linguistics (COLING 2012) pp.55-70 (2012.12)
8. Pascual Martínez-Gómez, Tadayoshi Hara and Akiko Aizawa:“Recognizing Personal Characteristics of Readers Using Eye-Movements and Text Features”, the 24th International Conference on Computational Linguistics (COLING 2012) pp.1747-1762 (2012.12)
9. Akihiro Kameda, Kiyoko Uchiyama, Hideaki Takeda, Akiko Aizawa: “Extraction of Semantic Relationships from Academic Papers using Syntactic Patterns”, eKNOW 2013, The Fifth International Conference on Information, Process, and Knowledge Management pp.32-35 (2013.02)
10. Towards a Data Hub for Biodiversity with LOD,MINAMI Yoshitaka, TAKEDA Hideaki, TAKEDA Hideaki, KATO Fumihiro, OHMUKAI Ikki, OHMUKAI Ikki, ARAI Noriko, JINBO Utsugi, ITO Motomi, KOBAYASHI Satoshi, KAWAMOTO Shoko,Lect Notes Comput Sci ,7774, 356-361,2013 年
11. Towards a Data Hub for Biodiversity with LOD,Yoshitaka Minami, Hideaki Takeda, Fumihiro Kato, Ikki Ohmukai, Noriko Arai, Utsugi Jinbo, Motomi Ito, Satoshi Kobayashi and Shoko Kawamoto,Proceedings of Joint International Semantic Technology Conference 2012 ,, 1-6, 2012 年 12 月
12. 情報モラル教育において抽象的概念を扱うための教授法の分析, 菅原真悟, 鷲林潤壱, 新井紀子, 日本教育工学会論文誌, 36(2), 135-146, 2012 年 10 月
13. 深い言語理解と数式処理の接合による入試数学問題解答システム, 松崎拓也, 岩根秀直, 穴井宏和, 穴井宏和, 相澤彰子, 新井紀子, 人工知能学会全国大会論文集(CD-ROM), 27th,,ROMBUNNO.2A4-1, 2013 年

〔著書等〕

1. Linked Data: Web をグローバルなデータ空間にする仕組み, トム ヒース, クリスチャン バイツァー(著), 武田英明(監訳), 大向一輝, 加藤文彦, 嘉村哲郎(, 亀田尭宙, 小出誠二, 深見嘉明, 松村冬子, 南佳孝(訳), 近代科学社, 2013 年 2 月
2. ほんとうにいいの? デジタル教科書 (岩波ブックレット), 新井紀子, 2012 年 12 月, 岩波書店

〔解説・総説〕

1. 岩根秀直, 松崎拓也, 穴井宏和, 穴井宏和, 新井紀子, 数式処理による入試数学問題の解法と言語処理との接合における課題, 人工知能学会全国大会論文集(CD-ROM) ,27th,, ROMBUNNO.2A4-2, 2013 年
2. 南佳孝, 武田英明, 加藤文彦, 大向一輝, 新井紀子, 神保宇嗣, 伊藤元己, 1 月博物館が所蔵する生物標本情報の Linked Open Data 化の試み,情報処理学会シンポジウム論文集, 2012,(7), 119-124, 2012 年 11 月
3. 稲邑哲也, 横野光, 新井紀子,物理モデル理解と自然言語処理の統合による試験問題の解答生成,日本ロボット学会学術講演会予稿集(CD−ROM) ,30th,,ROMBUNNO.3M2-1, 2012 年 9 月
4. 尾崎幸謙, 新井紀子, 土屋隆裕, 大学生数学基本調査の正答率補正(一般セッション 教育), 日本行動計量学会大会発表論文抄録集, 40, 329-330, 2012 年 9 月
5. 新井紀子,「合理的な思考」は計算可能か, 数学文化, 18, 2012 年 9 月

〔その他〕

1. H. Takeda: General Introduction for Semantic Web, Linked Open Data, in International Aasian Summer School on Linked Data, Daejeon, Korea (2012), Korean Advanced Institute of Scicence and Technology, Lecture.
2. H. Takeda: Identity and schema for Linked Open Data, in International Aasian Summer School on Linked Data, Daejeon, Korea (2012), Korean Advanced Institute of Scicence and Technology, Lecture.
3. H. Takeda and F. Kato: LOD Application Exemplar, in International Aasian Summer School on Linked Data, Daejeon, Korea (2012), Korean Advanced Institute of Scicence and Technology, Lecture.
4. 武田英明: Linked Data で社会と研究をつなぐ(<特集>編集委員今年の抱負 2013), 人工知能学会誌 28(1), 10, 2013-01-01 (2013)

＜会議発表等＞

〔招待講演・国内〕

1. 武田英明 : Linked Data における識別子とスキーマ, 特別講演, 第 28 回セマンティックウェブとオントロジー研究会, 人工知能学会 (2012).
2. 武田英明 : Linked Data がつくる新しいデータの世界, グリッド協議会　第 37 回ワークショップ　公共データのオープン化とクラウド, 東京.
3. 武田英明 : 識別子(ID)が作る新しい学術の世界, 第 6 回統合認証シンポジウム, 佐賀 (2012).

〔一般講演・国際〕

1. Akiko Aizawa, Michael Kohlhase, Iadh Ounis:“An Overview of NTCIR-10 Math Pilot Task”, MIR 2012 Workshop - Mathematics Information Retrieval at CICM 2012 (2012.06)

2. Giovanni Yoko Kristianto, Minh-Quoc Nghiem, Nobuo Inui, Goran Topic, Akiko Aizawa:“Annotating Mathematical Expression Definitions for Automatic Detection”, MIR 2012 Workshop - Mathematics Information Retrieval at CICM 2012 (2012.06)
3. F. Matsumura, I. Kobayashi, F. Kato, T. Kamura, I. Ohmukai and H. Takeda: Producing and Consuming Linked Open Data on Art with a Local Community, in J. F. Sequeda, A. Harth and O. Hartig eds., Proceedings of the Third International Workshop on Consuming Linked Data (COLD 2012) (2012), CEUR Workshop Proceedings Vol-905.
4. S. Koide and H. Takeda: Revisiting CLOS MOP, in International Lisp Conference, October 21-24, 2012, Miyako Messe, Kyoto, Japan, pp. 39–52 (2012).
5. R. Chawuthai, V. Wuwongse and H. Takeda: A Formal Approach to the Modelling of Digital Archives, in H.-H. Chen and G. Chowdhury eds., The Outreach of Digital Libraries: A Globalized Resource Network - 14th International Conference on Asia-Pacific Digital Libraries, ICADL 2012, Taipei, Taiwan, China, November 12-15, 2012. Proceedings, Vol. 7634 of Lecture Notes in Computer Science, pp. 179–188 (2012).
6. T. Goto, H. Takeda and M. Hamasaki: DashSearch LD: Exploratory Search for Linked Data, in Proceedings for the Second Joint International Semantic Technology Conference, JIST 2012, Nara, Japan, December (2012).
7. Y. Minami, H. Takeda1, F. Kato, I. Ohmukai, N. Arai, U. Jinbo, M. Ito, S. Kobayashi and S. Kawamoto: Towards a Data Hub for Biodiversity with LOD, in Proceedings for the Second Joint International Semantic Technology Conference, JIST 2012, Nara, Japan, December (2012).

〔一般講演・国内〕

1. Akiko Aizawa: “Reading as a translation process: issues in alignment of gaze and textual information”, 未来の翻訳研究に関するワークショップ (2012.12)
2. 相澤彰子:“視線情報と言語処理”, 情報処理学会東海支部第 7 回講演会 (2013.03)
3. 松村冬子, 嘉村哲郎, 加藤文彦, 小林巌生, 高橋徹, 上田洋, 大向一輝, 武田英明 : LODAC Museum: Linked Open Data による博物館情報の統合と活用, 人工知能学会全国大会(第 26 回)論文集, No. 3C2-OS-13b-9, 山口 (2012), 人工知能学会.
4. 神保宇嗣, 猪又敏男, 植村好延, 矢後勝也, 上田恭一郎, 南佳孝, 加藤文彦, 武田英明, 伊藤元己 : 種名データベース構築と情報の活用 : チョウ類を例に, 第 72 回日本昆虫学会大会, p. 45, 東京 (2012).
5. 武田英明 : DBpedia Japanese とは, in Wikimedia Conference Japan 2013, 東京 (2013).
6. 小出誠二, 武田英明 : 表示意味論にもとづく RDF/OWL 意味論の形式化, 第 29 回セマンティックウェブとオントロジー研究会, 東京 (2013), 人工知能学会.
7. 南佳孝, 武田英明, 加藤文彦, 新井紀子, 神保宇嗣, 伊藤元己 : 博物館が所蔵する生物標本情報の Linked Open Data 化の試み, じんもんこん 2012 論文集, 第 2012 巻, pp. 119 -- 124 情報処理学会 (2012).
8. 阿辺川武, 間下亜紀子, 文章中のコンテクストに適合した関連動画の検索. 情報処理学会研究報告エンタテインメントコンピューティング, 2013-EC-27(19), 2013.
9. 新井紀子,「Netcommons はどこへ向かうか」, Netcommons ユーザカンファレンス, 一橋講堂, 8 月 21 日
10. 新井紀子, Web 社会を生き抜く情報リテラシーをいかに育むか, Wikimedia Conference Japan 2013 2013 年 2 月 3 日

〔ポスター〕

［国　際］

1. R. Chawuthai, H. Takeda, V. Wuwongse and U. Jinbo: A Logical Model for Taxonomic Concepts for Expanding Knowledge using Linked Open Data, in Poster and Demonstration Proceedings, The Second Joint International Semantic Technology Conference, JIST 2012, Nara, Japan, December 2012, pp. 7–8 (2012).
2. F. Matsumura, F. Kato, T. Kamura, I. Ohmukai and H. Takeda: Generating LOD from Web: A Case Study on Building Integrated Museum Collection Data, in Poster and Demonstration Proceedings, The Second Joint International Semantic Technology Conference, JIST 2012, Nara, Japan, December 2012, pp. 23–24 (2012).

＜受　賞＞

1. 「生物情報基盤構築のための生物種データの Linked Open Data 化の試み」, 武田英明, 武田英明, 南佳孝, 加藤文彦, 大向一輝, 大向一輝, 新井紀子, 神保宇嗣, 伊藤元己, 小林悟志, 川本祥子, 人工知能学会全国大会優秀賞（口頭発表部門）, 2012 年 9 月
2. IARIA The Fifth International Conference on Information, Process, and Knowledge Management (eKNOW 2013) Best Paper Award "Extraction of Semantic Relationships from Academic Papers using Syntactic Patterns"

③　その他の成果発表

1. 高野明彦「蓄積から“創造”へ　～放送文化アーカイブ構想の可能性～」NHK 放送文化研究所春の研究発表とシンポジウム，パネリスト，2013 年 3 月
2. 阿辺川武「蓄積から“創造”へ　～放送文化アーカイブ構想の可能性～」NHK 放送文化研究所春の研究発表とシンポジウム，アーカイブサイト開発報告，2013 年 3 月
3. 阿辺川武，国立国会図書館 NDL ラボ「脚注表示機能を有した電子読書支援システムの構築実験」（http://lab.kn.ndl.go.jp/nii/），2013 年 5 月
4. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 12 名, 4 月 27 日,
5. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 7 名, 6 月 26 日
6. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 7 名, 8 月 24 日
7. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 9 名, 8 月 25 日
8. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 8 名, 10 月 12 日
9. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 4 名, 12 月 17 日
10. 新井紀子,「NetCommons 入門講座」国立情報学研究所, 9 名, 2 月 20 日
11. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 19 名, 5 月 29 日
12. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 5 名, 7 月 24 日
13. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 10 名, 9 月 10 日
14. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 5 名, 11 月 9 日
15. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 14 名, 1 月 24 日
16. 新井紀子,「NetCommons 活用講座」国立情報学研究所, 10 名, 3 月 21 日
17. 新井紀子,「NetCommmons デザインカスタマイズ講座」国立情報学研究所, 8 名, 6 月 25 日
18. 新井紀子,「NetCommmons デザインカスタマイズ講座」国立情報学研究所, 4 名, 9 月 11 日
19. 新井紀子,「NetCommmons デザインカスタマイズ講座」国立情報学研究所, 4 名, 12 月 18 日

20. 新井紀子,「NetCommmons デザインカスタマイズ講座」国立情報学研究所, 4 名, 3 月 22 日
21. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 6 名, 9 月 18 日
22. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 6 名, 10 月 10 日
23. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 5 名, 11 月 12 日
24. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 2 名, 12 月 19 日
25. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 4 名, 1 月 25 日
26. 新井紀子,「NetCommons モジュールカスタマイズ講座」国立情報学研究所, 3 名, 2 月 21 日
27. 新井紀子,「NetCommons システム運用講座」国立情報学研究所, 11 名, 6 月 22 日
28. 新井紀子,「NetCommons システム運用講座」国立情報学研究所, 3 名, 10 月 11 日
29. 新井紀子,「NetCommons システム運用講座」国立情報学研究所, 2 名, 1 月 28 日
30. 新井紀子「NetCommons システム運用講座」国立情報学研究所, 8 名, 2 月 15 日
31. 新井紀子, NetCommons ユーザカンファレンス 2012, 国立情報学研究所, 246 名, 8 月 7 日

＜新聞報道など＞

1. 新井紀子,「富士通、研究者業績情報を容易に公開できる大学向け SaaS,Ufinity, 研究者業績サービス,」, クラウド Watch, 2012/09/14
2. 新井紀子,「富士通、低コストで短期公開できる SaaS 型研究者業績サービスを開始」, 2012/09/14,,Computerworld