

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・ユーザインタフェース技術・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・オントロジーに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十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がある。

(3) 期待される成果等 (学問的效果、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

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・ユーザインタフェース技術・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 回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・分類手法に関する研究を進め、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、コンテキスト（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 Academia.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。

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

(2) 各年度の計画

平成 22 年度 (予備研究・プロジェクト開始)

サブテーマ「研究資源に関する情報推薦基盤の構築」の目標は、研究者プロフィール情報を活用した高性能な論文・研究者推薦エンジンの実現に向けて検討を行うことである。平成 21 年度に行った調査研究では、著書論文情報・引用情報に加えて「分野横断型融合プロジェクト (PD:東倉)」で抽出した研究者情報を相互に関連付け管理するデータベースを構築し、単語や論文の出現頻度に基づく類似度計算を可能にした。また、これらを利用した論文推薦システムのプロトタイプを試作して動作を確認した。今年度は、上記で作成したデータベースおよび学会誌論文のテキストを抽出した全文アーカイブを利用して論文推薦の実証的な実験を行うとともに、NII サービスの検索ログ、学術用語辞典、IEEE/Springer などの論文誌や会議録等の欧文全文データを情報推薦に利用するためのデータベース設計を行う。また、外部インタフェースを介して Researchmap をはじめとする共有基盤や論文検索サービスと連携して有用性を評価する。今後、さまざまな推薦手法を実装して性能を比較したり、論文推薦以外の検索システムを構築するためのプラットフォームとして活用し、研究を展開する予定である。

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」の目標は、学術情報を学術コミュニティおよび社会で共有できる仕組みを構築するものである。このうち、22 年度は基本技術の調査とデータ分析を主に行うとともにシステムの機能設計を行うための実験システムを作成する。具体的には(1)汎用の軽量オントロジーの構築、(2)セマンティックデータ交換形式の策定、および(3)軽量オントロジーを利用したアプリケーションのプロトタイプを作成の 3 点を行う。

サブテーマ「連想情報処理基盤の研究」の目標は、現在運営している文化遺産オンラインの文化財情報を起点にして、人物情報、論文情報、書籍情報、分野オントロジー、コースウェアなど、高信頼な公開情報を収集・整理する。収集された情報を対象に想・IMAGINE システムを立ち上げて、連想計算による結合と相互分類の実験を行う。また、1 つの情報源を軸として、他の情報源からの関連情報を集約して、新しい情報源を合成する方式についても検討する。さらに、このような連想計算による異種情報の結合・分類処理に適した情報の表現方法についても検討する。たとえば、コースウェアを軸に関連する人物や作品の情報を集約して、内容豊富な新しいコースウェアを合成する実験などを計画した。

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」の目標は、学術研究の多

様なデータおよびサービスを、各研究者に最適であるようにチューニングした上で提供することである、研究者向けのクラウドサービスを確立することと、サブテーマ(1)~(3)を有機的に接続させ、実サービスとして研究者コミュニティおよび社会に還元することにある。それには、(1)サブテーマで提案された各要素技術との連携(2)研究者コミュニティとの連携・データの取得(3)学術研究データベースおよび研究者サービスとの連携、が不可欠となる。今年度は、その準備段階として、第一期中期計画で構築した研究者向け情報共有基盤システム NetCommons および Researchmap β の改善を行い、Researchmap をサービスとしてリリースする。その上で、上記サブテーマに必要である、多様な学術・研究データの大規模取得、および、バーチャルな研究コミュニティへの支援を試みる。

サブテーマ1 「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」

研究者プロフィール情報を活用した高性能な論文推薦エンジンの実現に向けて検討を行う。著書論文情報・引用情報に加えて「分野横断型融合プロジェクト (PD:東倉)」で抽出した研究者情報を相互に関連付け管理するデータベースおよび学会誌論文のテキストを抽出した全文アーカイブを利用して論文推薦の実証的な実験を行うとともに、NII サービスの検索ログ、学術用語辞典、IEEE/Springer などの論文誌や会議録等の欧文全文データを情報推薦に利用するためのデータベース設計を行う。また、本サブテーマの目標である、各研究者にパーソナライズされた研究情報を的確に推薦するための手法の研究開発については、学術情報空間から各研究者に関する情報を自動的に収集することにより、共著関係や共同研究者などの実世界の間ネットワーク情報等を同定、分析し、研究者の研究上の興味関心を捉え得ることが重要であるため、Researchmap との連携に向けた API 設計を行う。

論文推薦のアルゴリズムについては、上記のパーソナライズされた情報だけでなく、論文の内容に基づいたコンテンツベースの手法を複数の視点から検討する。

サブテーマ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」

本研究ではデータ中心型研究基盤構築という全体目標達成のために以下の5つのトピックに関する研究を平行して実施する。

1. 基盤ソフトウェア環境構築
2. 基盤データベース構築
3. 研究コミュニティ支援サービス構築
4. 発展ソフトウェア開発
5. プラットフォーム構築

まず、本年度は基盤的なソフトウェア開発と基盤的データベース構築を中心として研究を進める。具体的には以下のことを行う。

1. 基盤ソフトウェア環境構築

1.1 Linked Data 基盤ソフトウェア整備

本プロジェクトではデータはすべて RDF(Resource Description Framework)で記述される。この RDF データを操作し、格納して、検索可能な環境を構築する。具体的にはソフトウェアとして利用可能な RDF ストア (RDF のデータベース) の性能を調査し、スケーラビリティ、スピード、推論、多言語対応 (日本語対応)、の視点から評価し、適切なものを選定する。また、RDF ストアの周辺環境 (Web UI など) は本プロジェクトのニーズに合わせて開発する。また、本プロジェクトが提供するデータの基礎的な構造 (オントロジー) を規定する。

1.2 スクレイピングソフトウェア開発

利用可能なデータがすでに RDF で公開されている場合は、単に RDF データの収集でよいが、

Web ページ等で公開されている場合は、Web ページから必要な情報だけを抜き出し（スクレイピング）、RDF データに変換する必要がある。このスクレイピングを行うソフトウェアを開発する。特に Deep Web（データベースに基づき Web ページが動的に生成されるような Web サイト）のスクレイピングが可能な仕組みを考える。なお、スクレイピングにおいては Web サイト側の負担に注意する。

2. 基盤データベース構築

データの共有は単にそれを可能とするソフトウェアの提供では不十分である。実際に利用可能なデータを提供することで初めてデータ共有の価値が認識される。このため本プロジェクトでは実際にいくつかの分野のデータを収集して利用可能にする。

2.1 美術館・博物館情報

国内には 5000 館を超える美術館・博物館が存在し、多種多様な美術品が収集公開されているが、その情報公開は限定されている上に、孤立している。このような状況は Linked Data アプローチによって解決可能である。実際にスクレイピングによりデータを収集し統合することで本アプローチに有効性を検証する。

サブテーマ 3 「連想情報処理基盤の研究」

本研究テーマでは、連想情報学における異種情報の結合・分類手法を深化させて、それぞれ独立に構築され公開されている多様な文化的情報源を動的に結合・融合・再構成する連想情報処理技術の確立を目指している。本研究では、博物館・美術館・図書館・文書館が歴史的に蓄えてきた情報の相互連携 (MLA 連携) を進め、それらを各種の問題解決に役立てるための情報環境の実現を目指す。そのためには、人物情報、論文情報、書籍情報、分野オントロジー、コースウェアなどの多種多様な情報の活用が不可欠であり、連想情報処理基盤はそのために欠かせない基本技術である。連想情報処理技術の確立により、背景知識や専門性が異なるユーザでも、自分が理解可能な情報を出発点として、さらに正確で信頼できる関連情報に辿り着き、それを学びながら情報探索を続けられる情報利用環境を構築できる。

これまでに我々が研究開発した世界最速の連想計算エンジン GETA によるコンテンツ・コンパイル技術は、蓄積された情報源の特徴を連想計算機構として抽出し、それを情報源同士の相互作用に活用する試みであり、世界的に見てもオリジナリティは高い。連想計算の応用方法の一つである連想検索技術に限定しても、数千万件規模のコンテンツに適用可能な類似技術は知られていない。本研究では、この技術をさらに深めて異種複数コンテンツの融合・再構成のための連想情報処理基盤を確立するが、これはコンテンツ間に”化学反応”を起こして”化合物”を作り出す試みと考えられる。

平成 22 年度は、従来、大学図書館蔵書情報中心の情報発信であった Webcat Plus を全面改変して、国立国会図書館の日本書誌目録、近代デジタルライブラリ、全国古書店在庫データベース、青空文庫、ウィキペディアなどを融合して、現存する和書全体をカバーするサービスとして新規開発を行う。さらに、それらの書誌データを、本だけでなく、作品や人物を加えた三つの軸で再整理して、それぞれについての検索機能を連携させる。また、選択した複数の本から関連本を連想検索する機能のために、ユーザが自分の興味のままに本やウィキペディア項目等を選択して保存できる「連想×書棚」を提供する。この仮想電子書棚は、連想検索の基点として役立つだけでなく、ユーザの興味を表現してシステムに伝えるためのメディアとして有効に働くと期待できる。Webcat Plus で作成した「書棚」を、Researchmap 等の外部サービスから利用できるプラグパーツを開発する。

図書館の連携として、国立国会図書館の次世代書籍検索サービスのプロトタイプシステム「国立国会図書館サーチ (試行版)」の構築に協力して、検索結果から連想計算により関連するウィキペディア項目を返す「連想キーワード」機能を提供する。

美術館との連携としては、徳川美術館、東京文化財研究所、慶応大学斯道文庫、小布施町における文化財情報の電子化に協力して、常設展や企画展での情報発信を分担する。

分野横断型の研究活動において本研究成果の有効性を評価する足がかりとして、「人類の移動誌」研究会を対象に、研究データや研究会資料の電子化に着手する。

Webcat Plus の全面改変により、現存する和書全体をカバーする国内最大の書誌データを研究利用できる情報環境が整った (<http://webcatplus.nii.ac.jp/>)。従来の大学図書館 1000 館の蔵書 DB に加え、国立国会図書館に納本されたすべての書籍を含む日本書誌目録、明治大正期の著作権切れ著作物の近代デジタルライブラリ、全国 850 店の古書店在庫データベース、目次情報を含むブックデータベース、著作権切れの作品 8 千点を全文電子化した青空文庫、ウィキペディアなどを融合して作成している。特に、ウィキペディア等から人物項目 14 万人を収集して、人物ごとに関連する作品や本を連想検索で自動収集する実験を行い、その有効性を確認した。同機能は新 Webcat Plus のサービスに組み込まれて一般公開されている。また、連想的対話技術の研究の一環として、ユーザの興味を表現して連想検索の基点となる情報を蓄えるための「連想×書棚」を開発して、新 Webcat Plus の公開サービスに組み込んだ。この「連想×書棚」は、予備知識の異なる異分野の専門家同士でのコミュニケーションにおいて、自分の興味を誤解されにくい形で表現して伝達するのに役立つ。この「連想×書棚」は、ユーザの興味をシステムに伝えるために有効なだけでなく、人間同士のコミュニケーションにも役立つと期待できる。その意味から、Webcat Plus で作成した「書棚」を Researchmap 等の外部サービスから利用できるブログパーツを開発した。

Webcat Plus におけるウィキペディア情報の活用として、Webcat Plus の書誌情報とウィキペディア項目を関連づける実験を行った。具体的には、連想検索結果の書籍群と関連するウィキペディア項目の収集、ウィキペディア項目語彙による特徴語抽出、ウィキペディアの項目間構造を用いた自動分類（カテゴリー付与）の実験を行い、それらの有効性を確認した。これらは Webcat Plus の公開サービスに組み込まれている。

国立国会図書館の次世代書籍検索サービス「国立国会図書館サーチ（試行版）」の構築に協力して、検索結果から連想計算により関連するウィキペディア項目を返す「連想キーワード」機能を提供した。徳川美術館、東京文化財研究所、慶応大学斯道文庫、小布施町における文化財情報の電子化に協力して、常設展や企画展での情報発信を分担した。

分野横断型の研究活動において本研究成果の有効性を評価する足がかりとして、「人類の移動誌」研究会を対象に、研究データや研究会資料の電子化に着手する。

サブテーマ 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」

本研究テーマでは、学術研究の多様なデータおよびサービスを、各研究者に最適であるようにチューニングした上で提供する、研究者向けのクラウドサービスを確立することと、サブテーマ(1)～(3)を有機的に接続させ、実サービスとして研究者コミュニティおよび社会に還元することが最終的な目標である。それには、(1)サブテーマで提案された各要素技術との連携(2)研究者コミュニティとの連携・データの取得(3)学術研究データベースおよび研究者サービスとの連携、が不可欠となる。今年度は、その準備段階として、第一期中期計画で構築した研究者向け情報共有基盤システム NetCommons および Researchmap β の改善を行い、Researchmap をサービスとしてリリースする。その上で、上記サブテーマに必要である、多様な学術・研究データの大規模取得、および、バーチャルな研究コミュニティへの支援を試みる。

平成 23 年度

サブテーマ 1 では、実験的に立ち上げたプロトタイプシステム上で論文推薦システムの課題や解決法を検討する。また、著者同定と引用文献同定に基づく個人プロフィールの自動獲得、引用および共著者・共同研究者ネットワーク情報の補間手法について検討する。前年度で設計したデータベースに、実際に運用に使われている論文データベースおよび著者情報を入力して、効率や性能の予備調査を行い、自動データ更新に向けて準備を進める。また、コンテンツベースの推薦性能を高めるため、論文主題スキーマの設計や論文主題の抽出などに向けた課題を整理する

サブテーマ 2 では、本年度は前年度から開始した基盤ソフトウェア環境を継続すると共に、この技術を使って研究データの取り扱いをするための準備を行う。前者については研究データとして散在する比較的小規模のテーブル型データを公開共有する仕組みについて取り込む。また後者としては具体的に研究コミュニティ（GBIF と GIS）と連携して、データベース構築と基盤ソフトウェアの改良を行う。

1. 基盤ソフトウェア環境構築

1.3 テーブル型データ取り込みソフトウェア整備

Web 公開されていないデータも多くは表計算ソフトウェアのデータ形式(excel 形式や csv 形式)で保存されていることが多い。この形式のデータを RDF 化して取り込むことはデータ活用において重要である。テーブル型データ取り込みソフトウェアはいくつかが公開されているが、それらを記述の容易性、表現力、多言語対応（日本語対応）の視点から評価して、必要なソフトウェアを選択する。そのうえで周辺ソフトウェアを開発して、簡便に利用可能な環境を構築する。

1.4 Researchmap ジオコーディング連携

地理情報は多くの研究分野で共通に使われるデータである。ジオコーディングではテキストデータに緯度経度といった地理情報を付加することでコンピュータが理解可能な情報にする。この仕組みが Researchmap で利用可能になるようなソフトウェアを開発する。単に緯度経度だけではなく地名情報も扱えるようにすることで、地図マッピング以外の利用を可能にする。

2. 基盤データベース構築

データの共有は単にそれを可能とするソフトウェアの提供では不十分である。実際に利用可能なデータを提供することで初めてデータ共有の価値が認識される。このため本プロジェクトでは実際にいくつかの分野のデータを収集して利用可能にする。

2.2 地理・地名情報

地理・地名情報は様々な分野で利用される基礎的データである。この情報を共有可能にすることは、単にその分野での情報活用に貢献するだけでなく、分野をまたがった情報共有を促進することに効果がある。本プロジェクトでは緯度経度といった地理情報だけでなく、住所情報、公共施設情報など地理に関わる様々な情報を Linked Data 化する。

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

環境プロジェクトでは分野が異なる多様な機関からのデータの統合的な取り扱いが必要である。これまで構築したデータベース、システムを利用して、特定の環境プロジェクトの活動を支援する仕組みを試行的に構築する。特に国立遺伝学研究所の GBIF (Global Biodiversity Information Facility) 活動と連携して、データの取り込み支援や他のデータとのリンケージなどの課題を解決する（共同研究者：菅原英明（遺伝研））。

3.2 GIS プロジェクト連携

地理情報は多様なデータを統合する鍵となるデータである。地理情報を中心に多様なデータをリ

リンクすることで、研究活動を支援する仕組みについて取り組む。国立極地研究所の GIS を利用して、他のデータとのリンケージを可能とする仕組みを試作して、研究者の研究活動が可視化できる仕組みの構築を行う（共同研究者：小林悟志（極地研））。

サブテーマ 3 では、連想情報処理基盤の研究については、文化遺産オンライン由来の文化財情報と、Webcat Plus 由来の人物情報（著者情報）を起点として、高信頼な公開情報を収集・整理する。文化財情報と人物情報の 2 つを軸として情報を集約整理する方式について検討する。

サブテーマ 4 では、ユーザインタフェイス等を検討した上で、以上の 3 つのサブテーマの研究成果を Researchmap に反映させ、研究者コミュニティに提供、融合研究を支援する。また、包括脳支援データベース等、他機関・他プロジェクトのデータベースとの連携を開始する。具体的には科学技術振興機構（JST）が提供している ReaD の基盤ソフトウェアとして Researchmap を提供する。ReaD には既に 20 万人以上の産学官の研究者が参加していることから、Researchmap は世界有数の研究融合基盤となる。また、JST と連携を深めることにより、JST の他のサービス（J-GLOBAL, J-RECIN 等）と連動することが可能となり、学術コミュニティだけでなく、産業界、また海外との連携が可能となる。また、収集した研究者情報を API で公開するための設計を行う。

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。研究者個人プロフィー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。

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ェア開発および基盤的データベースを完成させる。また、内外の研究コミュニティと連携して、データ共有、相互参照を可能とする仕組みを実現する。

1. 基盤ソフトウェア環境構築

1.5 データ統合ソフトウェア開発

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

2.3 地域発信情報

地域から発信される情報は様々な学術分野に活用可能な情報であるものの、構造化も共有化も不十分である。地域の NPO と協力して地域の状況に合わせて地域情報の Linked Data を促進させる。他の基盤データを利用することで地域を越えた連携を可能とする。

2.4 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 と各種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。

1. 研究コミュニティ支援サービス構築

1.1 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。前年度の試作に基づいて、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、可視化システムを実装する（共同研究者：菅原英明（遺伝研））。

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 情報を核とするデータ統合の仕組みを研究する。とくに Researchmap ジオコーディング連携（2-1）の結果と統合して、Researchmap 上で地理情報を含む多様なデータを記述できる仕組みを構築し、研究者の研究活動の可視化や研究者のコラボレーション支援を行う。（共同研究者：小林悟志（極地研））。

サブテーマ3では、連想情報処理基盤の研究については、引き続き文化遺産オンライン由来の文化財情報と、WebcatPlus 由来の人物情報を起点として、高信頼な公開情報を収集・整理する。文化財情報、人物情報、タイムラインなど3つ以上の軸を使って情報を集約整理する方式について検討する。

サブテーマ4では、前年度設計した API を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、大学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。論文の専門度や「手法」、「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。

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。

2. 発展ソフトウェア開発

2.1 インスタンスマッピングアルゴリズム開発

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。

2.2 データ活用アプリケーション

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。地理情報があるので、PC 版だけでなく、携帯できるモバイル版も開発する。

3. プラットフォーム構築

3.1 統合プラットフォーム開発

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 では、連想情報処理基盤の研究については、引き続き文化遺産オンライン由来の文化財情報と、Webcat Plus 由来の人物情報を起点として、高信頼な公開情報を収集・整理する。個々の文化財の高解像度写真、年表、歴史地図などを軸にして情報を集約整理して、それらに対する直感的な対話環境について検討する。

サブテーマ 4 では、前年に受けた中間評価を元に、要素技術を改良し、準備が整い次第、Researchmap 上に再統合する。この年度までに、各関連機関と連携して、各種学術研究データのライフサイクルの仕組みを完成させるとともに、実サービスとしての精度を高める。また、学術コミュニティだけでなく、産業界、また海外との連携を開始する。

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。

サブテーマ 2 では、本年度はより基盤システムに高度な機能を実装するための研究を行うと共にシステム連携を実現する。

4. 発展ソフトウェア開発

4.3 オントロジーマッピングアルゴリズム開発

異なるデータサイトからくるオントロジー情報（概念、カテゴリやスキーマに関する情報）は同じものを指していることがある。このオントロジーマッピングを実現するアルゴリズムを開発する。NII の市瀬らの提案するアルゴリズムを発展させ、実問題で適用できるようにし、実際にオントロジーマッピングを行い、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。

5. プラットフォーム構築

5.2 Researchmap 統合プラットフォーム開発

統合プラットフォームのアプリケーションとして Researchmap 上でデータ操作ができる環境を駆逐する。Researchmap 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。Researchmap 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。

サブテーマ 3 では、連想情報処理基盤の研究については、公開コースウェアを起点として、高信頼な公開情報を収集・整理する。収集された関連情報をコースウェアの参考情報として適切にひも付けるための技術的課題について明らかにする。コースウェアにおけるトピックの変遷を自動的に捕捉することにより、関連性の評価を調整する方式を検討する。

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では本年度は最終年度として基盤システムを完成させる。また発展的テーマとして、研究データをより広範に取り込む仕組みとして政論文からデータ抽出なども行う。

発展ソフトウェア開発

4.4 論文からのデータ抽出

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にする。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。

統合学術オントロジーの構築

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。

プラットフォーム構築

リポジトリ統合プラットフォーム開発

統合プラットフォームのアプリケーションとしてリポジトリシステム **WEKO** 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。

クラウド型データ中心型サービスの構築

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。

サブテーマ3では、連想情報処理基盤の研究については、これまで研究開発した要素技術を、文化遺産オンライン、**Webcat Plus**、**想・IMAGINE** などの公開サービスにフィードバックして、それらの新しいサービス構築に活用する。

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。

サブテーマ2では、データに基づくバーチャル研究環境の構築

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。このために、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。

サブテーマ3では、世界最速の連想計算エンジン GETA によるコンテンツ・コンパイル技術が、我が国を代表する博物館・美術館・図書館・文書館で蓄積されている高信頼な情報源に適用され、それらが我々ユーザの日常的な意思決定に活用できる情報サービスとして提供される。本研究の技術をさらに深めて異種複数コンテンツの融合・再構成のための連想情報処理基盤が確立されれば、コンテンツ間に“化学反応”を起こして“化合物”を作り出す情報環境を実現できると考えられる。

[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

〔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

〔国立遺伝学研究所〕 菅原英明

〔東京芸術大学〕 嘉村哲郎

(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〕 西岡真吾、丸川雄三、阿辺川武、特任研究員 2名

〔国立遺伝学研究所〕 大久保公策

(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〕 羽田昭裕、山地一禎

〔情報・システム研究機構〕 舛川竜治、南 佳孝

〔国立極地研究所〕 岡田雅樹、小林悟志

〔総合研究大学院大学〕 大田竜也

〔電気通信大学〕 Neil Rubens

〔藤田保健衛生大学〕 宮川 剛

〔情報通信研究機構〕 内山将夫

〔東京大学〕 前田邦弘（クオンタム・アイディ）

〔4〕 研究の進捗状況

サブテーマ1

研究者プロフィール情報を活用した高性能な論文推薦エンジンのプロトタイプ（以下、「オススメ論文検索システム」と呼ぶ）の開発に着手し、必要となる要素技術について検討しながら予備的な利用者実験を行った。オススメ論文検索システムを実装する上で、(1)システム設計、(2)パーソナライズされた研究者情報に基づく推薦、(3)論文を複数の視点から推薦の3つが大きな柱となっている。この3つを中心に、研究の進捗状況、特に本年度実施した内容と、構築したシステムを使った評価実験の結果について報告する。

システム設計

まず(1)のシステム設計においては、データベースの構築、インタフェースの設計を行った。データベースは、CiNii 収録学会誌論文の抄録データ、学会誌論文（情報処理学会）のテキストを抽出した全文アーカイブ、IEEE/CS、Springer、Kluwer、Oxford University Press などに収録されている論文誌や会議録等の欧文抄録データに基づいて構築した。インタフェースの設計では、研究者情報に基づく推薦を可能にするために、著者名入力することによりログインする方法だけでなく、キーワード入力による論文推薦を可能にする方法も並行して実装した。これは、過去に論文執筆の経験がない学部学生などの初心者や、自分の分野とは異なる新しい分野の動向を知りたい研究者用として設計した。図1に操作の流れを示す。

システム操作の流れとしては、入力（キーワードあるいは研究者名）に対してその入力に関連した論文をリストに表示し、リストの中からシードとなる論文を選択する。シード論文を選択することにより、8つの観点からの推薦論文リストが表示され、その中から興味のある論文の詳細を CiNii で確認するというのが一連の流れである。



図 1: 推薦システムの操作の流れ

パーソナライズされた研究者情報に基づく推薦

パーソナライズされた研究者情報に基づく推薦については、著者名を入力することにより、著者が過去に執筆した論文リストを表示し、その執筆論文のキーワードから類似した論文および類似した研究を実施している研究者のリストも表示する。検索ログや研究者プロフィールがない状態における協調フィルタリングを実現するための基礎的な手法である。システムに論文推薦のログを保存する機能を付与し、ユーザが入力したキーワード、選択したシード論文、閲覧した論文等のログ取得が可能になった。また、Researchmap と研究者 ID の受け渡しのインタフェースを実装した。今後 Researchmap との連携により、研究者プロフィールや研究者ネットワークなどの情報入手が可能になれば、それらの情報を利用した協調フィルタリング手法を用いてパーソナライズされた情報に基づく推薦を検討していく予定である。

複数の視点から推薦

複数の視点から推薦する手法の検討は、本年度の中心的な課題であった。まず、システムを利用する対象を設定することにより、推薦する観点を絞り込んでいった。当初は、対象者を研究者全てと漠然としていたが、論文を検索するのに難しさを感じるのは、研究をこれから始める初心者である考えた。ある特定分野についてこれから学ぼうとする学生や、分野の概要を理解したい他分野の研究者などを対象に設定した。初心者が調べたい内容、欲しい情報を効率よく表示できるように必要な観点を以下の 8 つに整理した。

1. 類似論文レコメンダ
2. 国際論文レコメンダ
3. 速報論文レコメンダ
4. 発掘論文レコメンダ
5. 対象論文レコメンダ
6. 手法論文レコメンダ
7. 入門論文レコメンダ
8. 基礎論文レコメンダ

(1) 類似論文レコメンダ

シード論文に選んだ論文と類似した論文を推薦する。書誌情報のタイトルおよび抄録に含まれる特徴語 (Unidic 辞書の名詞に該当する語) を用いて、ベクトル (tf·idf) を算出し、シード論文の特徴語ベクトルと類似した論文をコサイン類似度で計算し上位 100 位の論文から 3 論文を表示する。疑似論文レコメンダは、表示された論文からシード論文に類似した興味ある論文や、先行研究の調査として類似論

文の傾向を調べる時などに役立つレコメンダである。

(2) 国際論文レコメンダ

英語で発表された論文を推薦する。日本語の論文においても、特に工学系の論文では、英語のタイトル、抄録、キーワードを記述することが義務付けられている。この情報を利用して、利用者が日本語の論文を選択した場合でも、その日本語の論文に含まれる英語の情報を手がかりにして、英語の論文を推薦する仕組みを構築した。国際論文は、NIIの電子情報リポジトリ NII-REO に収録されている論文誌を対象として検索しています。このレコメンダを利用することにより、日本語の論文から関連する英語論文を検索することが容易となり、海外の研究動向を把握するのに役立つ。特に、初心者の場合、日本語のキーワードを知っていてもその訳が分からないことが多いことから、重要なレコメンダであると考えられる。ただ、現在のアルゴリズムでは、日本語と英語の両方の情報を含む論文でないと国際論文を推薦できないため、日本語だけしかない論文から多言語の論文を推薦するアルゴリズムについて現在検討している。

(3) 速報論文レコメンダ

類似論文レコメンダで検索した論文 200 件を、新しく発表された順に表示するレコメンダである。分野の最新動向を調べる時に便利なレコメンダである。

(4) 発掘レコメンダ

異なる分野で発表された類似論文を推薦する。著者やその共著者が投稿しない論文誌を異なる分野と考え、馴染みのある論文誌以外に類似した内容の論文が掲載されていないかを検索する。このレコメンダでは、トップダウン式に分野を分類するのではなく、著者が投稿するかしないかという情報に基づいて分野の判定を行っている。

(5) 対象レコメンダ

シード論文のタイトルから、対象を表す専門用語を抽出し、対象が共通の論文を推薦する。たとえば、手がかり語を間に挟んだ「A による B」「A のための B」等の場合、B に該当する用語が対象を表す用語とする。ここでは、対象を表す手がかり語を 23 種類用意している。対象が共通している論文に絞って効率的に検索することが可能である。現在は、タイトルに含まれる用語だけを抽出していることから、推薦可能な論文が限られてしまうことから、抄録中からも用語を抽出する方針を検討している。

(6) 手法レコメンダ

シード論文のタイトルから、手法を表す専門用語を抽出し、手法が共通する論文を推薦する。対象レコメンダと同様で、手がかり語を間に挟んだ「A に基づく B」「A を利用した B」などの表現の場合、A が手法を表す用語とする。手がかり語は 37 種類用意し、同じ手法を用いた論文を効率的に推薦できる。

(7) 入門レコメンダ

初心者が最初に読むべき基礎的な論文や、解説記事を優先的に推薦する。タイトルや抄録に含まれる複合語を、分野における重要性・基礎性が高い用語から、分野特有の難解な用語まで度合いを分類し、その分野では必須の用語を多く含む論文を推薦する仕組みを取り入れている。論文の読みやすさなども重要であるが、現時点では、用語の難易度に従って論文を推薦するようにしている。今後は、用語の基礎性に基づいて、論文全体としての読みやすさ、基礎性について検討していく。

(8) 基礎論文レコメンダ

関連する論文によく引用されている基礎的な論文を推薦する。論文の中で引用している引用論文情報を用いて、類似論文中で被引用されている基礎的な論文を表示する。多く引用されている論文は、重要な論文であると定義し、分野において読んでおくべき基礎論文を検索することが可能となる。引用文献情報が限定されているため、推薦できる論文がないことが多く、類似論文だけでなく、その分野において引用数が多い論文を推薦できるように検討する必要がある。

システムインタフェース

以上、8つの観点（レコメンダ）による推薦を表示した画面を図2に示す。この画面は、キーワード入力の結果、関連論文リストの中からシードとなる論文の一つを選択し、その論文情報を上方に表示し、下方に各レコメンダが推薦する論文を表示している。

オススメ論文検索 

[ログイン](#) [\[管理専用\]](#)

図 2: オススメ論文検索システムの推薦結果画面

システムの実験

次に被験者がオススメ論文検索システムを使って、興味を持っているテーマについて情報推薦機能を評価してもらった。具体的には、被験者が興味のあるテーマに関するキーワードを入力し、そのキーワ

ードに関連する論文リストの中から、最も興味のある論文を選択（キー論文）し、その論文をキーとして8つのリコメンダがそれぞれ3つの論文を推薦し、その結果を個別評価する。個別評価は5回繰り返す、その後全体評価を実施する。

評価内容

評価内容は以下の通り、個別評価と全体評価を行った。なお、実験の最後に被験者の執筆論文数、研究年数、研究分野を記入してもらった。

(ア) 個別評価

- 速報リコメンダ：興味ある最新の論文が推薦されていましたか？
- 類似論文リコメンダ：選択した論文と類似した論文が推薦されていましたか？
- 国際論文リコメンダ：興味ある英語論文が推薦されていましたか？
- 発掘リコメンダ：異なる分野に興味のある論文が推薦されていましたか？
- 対象リコメンダ：対象（目的）が同じ論文が推薦されていましたか？
- 手法リコメンダ：手法が共通した論文が推薦されていましたか？
- 入門リコメンダ：初心者向けのわかりやすい論文が推薦されていましたか？
- 基礎論文リコメンダ：信頼性の高い論文が推薦されていましたか？

(イ) 全体評価

- 全体評価としてシステムを使った感想、意見等を記入
- 気に入ったリコメンダを選択（複数選択可）
- 気に入ったリコメンダについてその理由を記入

評価

評価結果の概要は以下の通りである。個別評価、全体評価を集計し、その結果を表1から表3にまとめた。

被験者数：35名

被験者の条件：論文（全国大会を含む）を執筆した経験を持つ人

被験者の平均執筆論文数：10本

被験者の平均研究年数：10年

被験者の研究分野：工学、自然科学、医学、言語学、文学、医学、生物学

システムの問題点・意見について

表1は全体評価のシステムを使った意見の代表的なものをまとめて、それに対する対応案について記述したものである。被験者の意見として、検索方法（キーワード補足）、検索結果の精度向上（論文のクラスタリング）、システムインタフェース（付加的表示機能）など有益なものであった。自然言語処理や情報システム関連の研究者については、被験者の感想については、肯定的なものが13件、否定的なものが9件、肯定的・否定的両方含むものが9件、無回答が4件であった。この結果から概ね全体的な印象としてシステムに対して高評価を得たと考えられる。

被験者の研究分野で目立った印象は、生物学（ウイルス学、植物細胞分子生物学）の被験者には完全にシステムを否定された。理由としてCiNiiに収録されている論文は工学系が多いため、あまりヒットする論文が少なかった点と、本システムが日本語の論文推薦に重きを置いているため、国際論文を重視する分野の研究者には評判が悪かった。また、自然言語処理、情報工学の研究者には推薦した論文の精

度に対して否定的な意見が多かった。検索の仕組みを知っているため、評価も厳しくなったと考えられる。

表 1: 被験者からの主な問題点とそれに対する対応案

問題点・提案	対 応
キーワードの補足	オントロジーの利用
キーワードの絞り込みが難しい	
ユーザによるリコメンダの選択	必要
論文を分野別にクラスタリングして、異なる分野論文の表示を制限する	分野クラスタリングをやるべき、発掘にも適用
類似と発掘が同じ論文	
キー論文の内容が概説、手法なのかを判断	分野基礎性をうまく組み込めば可能
国際ソート機能（新着順、インパクトファクター順）	あった方がよい
「表示件数/全体数」の表示	閾値を設定する？
検索キーワードとの類似度の数値表示	
検索が遅い	個人差、環境にもよるのでは？
同じ著者の論文はまとめる	検討予定
古典リコメンダを作る、それ以外は新しいもの	
抄録の一部の表示	

推薦論文の評価

レコメンダが推薦した論文が適切であったかどうかを 5 段階評価で評価をしてもらった結果を表 2 にまとめた。類似論文が最も評価が高く、普通よりも良いと判断されていた。他のリコメンダはあまり評価が高くなく、基礎、手法、対象については推薦される論文がないことが多かったため、全体的に評価が非常に低くなっている。

表 2: 推薦論文の適切性の評価

リコメンダ	平均評価値
類 似	3.46
発 掘	2.93
速 報	2.88
国 際	2.84
入 門	2.59
基 礎	0.6
手 法	0.45
対 象	0.03

レコメンダの評価

気に入ったリコメンダを複数記述してもらい集計した結果を表 3 にまとめた。類似リコメンダは推薦

論文として適切であったという評価も影響して、一番人気であった。次が入門リコメンダであった。入門リコメンダは推薦論文の適切性は5番目と非常に低い評価であったが、人気は二番目という結果となった。理由として、自分の分野については論文の検索方法がわかっているが、新しい分野については何から調べて読めばよいのかわからず、入門リコメンダがあると非常に便利であるという意見であった。入門リコメンダについては推薦論文の質が悪かったが、視点としては有益であるとの評価を得た。その次が国際論文リコメンダであるが、これは分野によって差が出ており、普段あまり国際論文をチェックしない日本語学の研究者には評判が良かった。

表 3: 被験者が気に入ったリコメンダと被験者数

リコメンダの種類	気に入った被験者数
類 似	18
入 門	12
国 際	8
速 報	7
発 掘	6
基 礎	5
手 法	1
対 象	0

考察

全体的に高評価であったが、個々の推薦論文の質に対しては不満が残る結果であった。多視点によるリコメンダの発想はおもしろいが、そこから推薦される論文はリコメンダの特徴が出ていないものが多かったためである。8つの視点は数が多く、重複して推薦される論文も多かったため、リコメンダの数を減らし、各特徴をうまくいかせるような仕組みを考えていく必要がある。

サブテーマ 2

本年度は基盤的ソフトウェア整備と基盤的データベース整備を中心に活動を行った。

本プロジェクトでは、広く学術に関する情報・データを共有する仕組みを **Linked Data** の方法で構築するということを目指している。このプロジェクトでは単に情報基盤を構築して分野の研究者やコミュニティに供給するというのではなく、当該分野の研究者と一緒に使って実際に使える形のサービスを作るところまでを実験的に行うことを狙っている。

Linked Data のメリット

Linked Data を用いることのメリットは3段階に分けることができる。

(1) オープンデータ化の手段

Linked Data はデータを公開して共有する手段として有効である。**URI** を識別子として使うことで、データ内の個別の事項をグローバルに一意的に指すことができる。また標準的な方法でアクセスすることができる。またデータの表現が **RDF** という一つの方法であるので、データ処理において単純化される。

Web でデータ公開をしていると言っても実際には **HTML** 文書、**PDF** 文書、**csv** など様々な方法がとられている。これらに比べ、**Linked Data** によるデータ公開はアクセスの容易性やデータの再利用性で優れている。

(2) 分野内データ共有の手段

一つの分野内でもデータを公開しているサイトは一つであるとうことは稀であろう。むしろ、複数のサイトが各々のもつデータを公開するというのが普通である。このような分散的な情報を統合的に扱うときにも **Linked Data** は効果がある。

(1)で述べたようにデータのアクセス手段・表現手段が統一されていることに加え、データサイトが相互にリンクを張ることにより関係性も表現できる。またデータ構造を規定するスキーマの対応も表現できるので、類似のデータ構造であれば、その違いを吸収してアクセスすることも可能である。

データをどこかに集約するのではなく、分散されたデータを分散しつつ、統合的なアクセス・利用を可能にするという点で **Linked Data** は有効な方法である。

(3) 分野を超えたデータ共有

さらに分野を超えたデータ利用にも **Linked Data** は効果がある。一般に他分野のデータはアクセス手段やデータ構造がわからず利用が難しい。**Linked Data** であればアクセス手段は一定のものが保証される上、公開されたスキーマはオントロジーとして表現することができるので、他分野のスキーマを理解して利用することが容易になることが期待できる。

分野を超えたデータ利用はデータに新しい価値を付加する。これがまたデータを公開することのメリットとなる。

3. 美術館・博物館情報

まず最初に注目したのは美術館・博物館情報である。美術館・博物館の情報は一部の集約サイト（文化遺産オンライン <http://bunka.nii.ac.jp/>、国立美術館所蔵作品総合目録検索システム <http://search.artmuseums.go.jp/>）をのぞけば、館ごとに情報公開がなされており、きわめて分散的である。従って(2)のメリットは期待できる。**LODAC Museum** と名付けたこのプロジェクトではこのような分散された情報を統合的に利用する技術を開発するためのプロタイプである。

3.1 **LODAC Museum** の情報源

日本には 6000 館に近い美術館・博物館が存在するが、Web により収蔵品や展示品の情報を公開している館はそれほど多くない。また公開している場合でも HTML ページとして公開しており、データの共有や利用においては問題がある。

本来は情報を公開する館が **Linked Data** として公開するのが望ましいが、今回は我々がデータを **scraping** し、**Linked Data** に変換している。現在、14 館の情報を収集している。他にも国指定文化財データベースなども用いている。また Wikipedia を簡易的に **Linked Data** 化する **dbpedialite** を通じて日本語 Wikipedia の情報も利用している。

3.2 データの表現と統合の方法

このような複数の情報源からの情報を扱うと、どう統合するかが問題になる。ここでは次のような方針を採った。

(1) 標準的なスキーマによるデータ表現

美術品のメタデータは詳細度の程度の差があるものの、表現内容はそう大きく変わるものではない。そこで、今回対象にした館での美術品のメタデータをおおよそカバーする程度のメタデータを定義し、個々の館からのデータをこれに当てはめた。

(2) 情報源ごとの ID と統合 ID

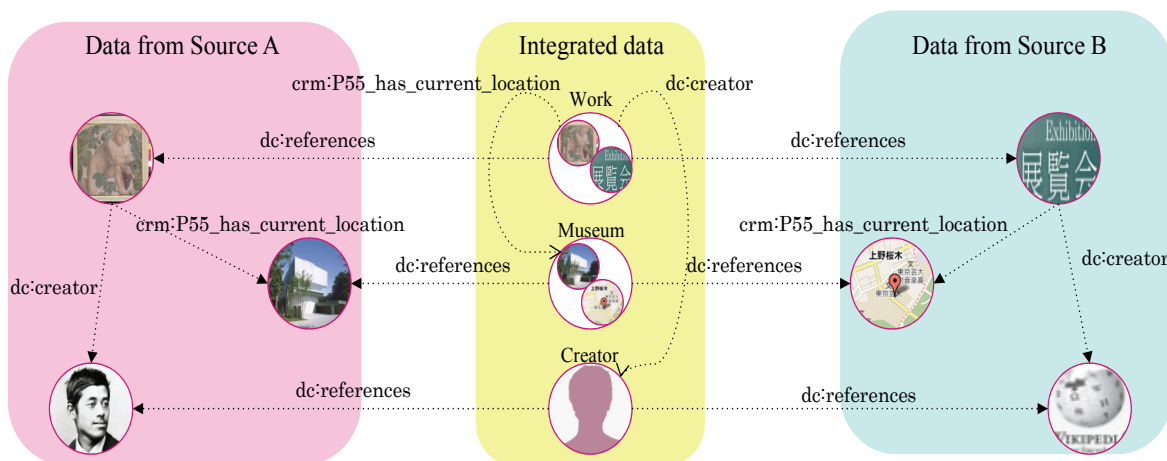
個別の館からのデータ（作品、作者）にはそれぞれに ID をつける。一方、それとは独立に統合 ID を個々のデータに振る。もし、一つの作品データや作者データが二つ以上の情報源からくる場合は、一つの統合 ID に複数の情報源毎ごとの ID が関係づけられられる（図 1 参照）。

(3) 統合のためのシソーラス

上記のような統合をするときには信頼できるシソーラスが重要になる。今回は日本美術シソーラス[福田 1997]を利用して作者情報の統合を行った。

3.3 実装状況

RDF データベースとしては BigOWLIM を使い、SPARQL Endpoint (<http://lod.ac/sparql>) と簡易ブラウジングインタフェース (<http://lod.ac/>) を用意している。現在、約 10 万のデータ (作品、作者等) が格納されている。上記の方法による統合結果としては、日本美術シソーラスにある 1332 人のうち 615 人が作品データに使われ、作品数の約 1/4(15020/61861)をカバーしている



28

図 3: 統合における ID の扱い

4. その他の分野

これ以外の分野のデータについては試験的に取り込みデータベース化を始めた。

4.1 地理・地図情報

地理・地図情報はデータを共有する軸として良く使われる情報の一つである。LODAC Museum には美術館・でも博物館施設の情報が含まれている。それを他分野のデータとも容易に結び付けられるようにするために、より汎用的な地理・地図情報の Linked Data 化を試みている。

最初に注目したのは住所情報である。住所情報として、(大字・町丁目レベル位置参照情報 (<http://nlftp.mlit.go.jp/isj/http://nlftp.mlit.go.jp/isj/>) と、郵便番号データ (<http://www.post.japanpost.jp/zipcode/>) の Linked Data、<http://www.post.japanpost.jp/zipcode/>) の Linked Data を試作した。また、汎用の施設情報として、国土数値情報 施設情報 (<http://nlftp.mlit.go.jp/ksj/http://nlftp.mlit.go.jp/ksj/>) と駅データ (<http://www.ekidata.jp/>) を Linked Data 化して、それを住所情報に結びつけることも試行している。現在 <http://location.lod.ac/> にて試験公開をしているが。

4.2 イベント情報

分野を超えた横断的データ利用の実験として、地域 NPO と協力して、地域イベントの Linked Data 化も試行している。イベント情報には地理情報、出演者・出展者といった人物情報が含まれ、他のデータとリンクされて利用される。

5. まとめ

本年度は基盤ソフトウェア開発と基盤的データベース構築を行った。データとしては美術館・博物館データをあつかったが、この処理において大規模データの取り扱いにおける課題や緩やかな分散デ

ータ統合などの課題が同定され、その解決を図った。これらの課題は研究データにおいても生じる共通の課題であり、その解決はデータ中心型研究基盤構築に大きく貢献する。今後は、今年度開発してきたデータ取り扱い手法を研究データに適用していく予定である。

サブテーマ 3

Webcat Plus の全面改変により、現存する和書全体をカバーする国内最大の書誌データを研究利用できる情報環境が整った (<http://webcatplus.nii.ac.jp/>)。従来の大学図書館 1000 館の蔵書 DB に加え、国立国会図書館に納本されたすべての書籍を含む日本書誌目録、明治大正期の著作権切れ著作物の近代デジタルライブラリ、全国 850 店の古書店在庫データベース、目次情報を含むブックデータベース、著作権切れの作品 8 千点を全文電子化した青空文庫、ウィキペディアなどを融合して作成している。特に、ウィキペディア等から人物項目 14 万人を収集して、人物ごとに関連する作品や本を連想検索で自動収集する実験を行い、その有効性を確認した。同機能は新 Webcat Plus のサービスに組み込まれて一般公開されている。また、連想的対話技術の研究の一環として、ユーザの興味を表現して連想検索の基点となる情報を蓄えるための「連想×書棚」を開発して、新 Webcat Plus の公開サービスに組み込んだ。この「連想×書棚」は、予備知識の異なる異分野の専門家同士でのコミュニケーションにおいて、自分の興味を誤解されにくい形で表現して伝達するのに役立つ。この「連想×書棚」は、ユーザの興味をシステムに伝えるために有効だけでなく、人間同士のコミュニケーションにも役立つと期待できる。その意味から、Webcat Plus で作成した「書棚」を Researchmap 等の外部サービスから利用できるブログパーツを開発した。



図 4: WebcatPlus のブログパーツ機能



図 5: WebcatPlus の RSS 配信機能

Webcat Plus におけるウィキペディア情報の活用として、Webcat Plus の書誌情報とウィキペディア項目を関連づける実験を行った。具体的には、連想検索結果の書籍群と関連するウィキペディア項目の収集、ウィキペディア項目語彙による特徴語抽出、ウィキペディアの項目間構造を用いた自動分類（カテゴリー付与）の実験を行い、それらの有効性を確認した。これらは Webcat Plus の公開サービスに組み込まれている。

国立国会図書館の次世代書籍検索サービス「国立国会図書館サーチ（試行版）」の構築に協力して、検索結果から連想計算により関連するウィキペディア項目を返す「連想キーワード」機能を提供した。徳川美術館、東京文化財研究所、慶応大学斯道文庫、小布施町における文化財情報の電子化に協力して、常設展や企画展での情報発信を分担した。

分野横断型の研究活動において本研究成果の有効性を評価する足がかりとして、「人類の移動誌」研究会を対象に、研究データや研究会資料の電子化に着手する。

サブテーマ 4

第一期中期計画で構築した研究者向け情報共有基盤システム NetCommons および Researchmap β の改善を行い、Researchmap をサービスとして正式にリリースした。

Researchmap は研究者に対して、研究ホームページを公開するための領域である「マイポータル」のほか、バーチャルなデスクトップの機能を果たす「マイルーム」、他の研究者と共同研究や委員会活動をするためのコミュニティを提供する。マイポータルには研究者履歴（Curriculum Vitae）を公開するためのテンプレートのほか、研究ブログ、資料配布用キャビネット、動画配信ツールなどが備えられており、研究者はその中から自分を表現するためのツールを自由にチョイスし、効果的に情報発信を行うことができる。

(1) マイポータル

ID を持つ研究者が Researchmap にログインしたときに最初に表示されるのがマイポータルである。マイポータルは、研究者個人が自分のためのホームページを公開するための領域として提供される。

その最初のページには、CV（Curriculum Vitae）モジュールがあらかじめ用意されており、研究者はここに自分の経歴や論文リスト、研究領域や研究キーワード等を登録・公開することができる。



図 6: RESEARCHMAP の CV モジュール

Researchmap では、CiNii, Pubmed, ArXiv, CiteSeer, KAKEN, Amazon 等から API 経由で論文情報・書誌情報・経歴情報・外部研究資金獲得状況などを取得することで手入力を経ずに CV を半自動生成する機能をもつ。また、生成した CV 情報をテキスト、CSV 等の形式でエクスポートすることが可能である。これにより、研究者情報の手入力の手間が省かれるだけでなく、論文や書籍の所在データが研究者の情報にリンクされ、研究情報がエンリッチされる。重層的な情報は、コンピュータが研究コミュニティ全体の動向を解析する上での貴重な付加的情報となる。これにより、爆発的に増殖している研究情報上の情報検索・情報推薦・情報可視化の精度を向上させることが可能となる。CV の各項目は「一般公開」「研究者のみに公開」「非公開」の 3 つのレベルからアクセスレベルを選択することができるため、「所属と論文リストは一般公開するが、経歴とメールアドレスは研究者のみに公開する」というようにユーザごとに自由に設定することができる。

平成 23 年度には、収集した CV 情報を他システムに提供するための API を設計し、平成 24 年度には実際の提供を開始し、研究情報の「エコシステム」を循環させる。そのテストケースとして、情報・システム研究機構が平成 22 年度に構築した「羽ばたけ日本の女性研究者」への API 提供し、自動的に女性研究者総覧の実験を行う予定である。

研究者情報の集約、API 連携、可視化は現在国際的にもホットな話題であり、米国では科学技術政策に対してより効率的に投資することを目的とした「StaResearchmapetrics」の一環として、コンピュータに解読可能であるような形式の下での研究情報公開のための基盤開発の必要性が説かれている。商用サービスでは、ResearcherID 等が研究者に特化したサービスを開始しているが、中立性に乏しい点が問題となっている。比較すると Researchmap は国際的にも最先端を走っているといえる。

Researchmap では、CV 以外にも、研究ブログを公開するためのツールや、講義資料等を配信するための資料庫、写真をスライドショーで公開するためのフォトアルバムなども提供している。どのモジュールについてもコメント等の双方向の仕組みが備えられているが、記入できるのは RESEARCHMAP に ID をもつ研究者に限られる。さらに、特定の研究者からのコメントだけを制限する仕組みも備えられている。平成 22 年度には、南極観測隊隊員である小林悟志研究員による南極観測ブログの発信を試み、新領域融合研究プロジェクト「地球環境変動の解析と地球生命システム学の構築」で行っている、苔坊主の採集の様などを発信し、サイエンスコミュニケーションの推進に役立てることができた。

researchmap 日本語 | English 新規登録 | ログイン

ホーム 研究者検索 コミュニティ検索

小林悟志

マイポータル
研究ブログ
資料公開

おとなりの研究者 | 注目の...

NO IMAGE 上野真義 12/22 更新	NO IMAGE 吉丸博志 11/27 更新
渡邊和男 11/25 更新	浜地 真志 05/06 更新
NO IMAGE 土畑重人 04/01 更新	NO IMAGE 津村義彦 12/21 更新

more..

カウンタ
COUNTER 86242

研究ブログ

研究ブログ >> 記事詳細 < 前の記事へ 次の記事へ >

2011/04/18 コケ 坊主の処理 by:iwana

2月19日。しらせ艦内の第2観測室にて、コケ 坊主の処理を行う。
南極の観測は、しらせに戻っておしまいではない。

帰路のしらせの中でも、観測物資の整理、サンプルの後処理と整理。
そして報告書の作成。。。
なかなか忙しい。

今日は、採取したコケ 坊主の整理。
湯の水に入れて持ち運んだコケ 坊主を形が崩れないように注意深く取り出し、ビニール袋に入れ冷凍室へ。
そして、コケ 坊主から顕微鏡観察用にチューブに分けて整理する。
残りのコケ 坊主は、形が崩れていないかを確認して冷凍室に入れた。

コケ 坊主の表面はコケに覆がまわりついている状態になっている。
コケの種類はLeptobryum(レプトブリウム)。もう一種類、Bryum(ブリウム)は今の段階では確認出来なかったが、研究室に持ち帰ってから、詳しく調べる予定。



(採取したコケ 坊主)



(顕微鏡に見えるのがコケ)

図 7:「地球環境変動の解析と地球生命システム学の構築」プロジェクトの広報ブログとしての利用

(2) マイルーム

ID を持つ研究者は、情報公開用のマイポータル以外に、自分自身のみがアクセス可能なプライベートスペースであるマイルームを取得できる。マイルームは研究者の「バーチャルデスクトップ」として位置付けることができよう。

その最初のページには、Web メール同様の機能を持つ PM (Private Message) モジュールがあらかじめ用意されている。これによって、ID を持つ研究者は互いに連絡を取り合うことができるような仕組みを実現している。

Researchmap の PM モジュールでは、画像やファイルを文書の指定した場所に HTML のように埋め込むことが可能である。これにより、よりデータ等を含む研究上の複雑なメッセージをより理解しやすい形で伝えることができる。また、メッセージごとにタグをつけて保存することで、整理しやすい構造となっている。



図 8: Researchmap の PM(プライベートメッセージ)モジュール

PM モジュールで届いたメッセージは通常のメールアドレスに転送することもできる。

マイポータルと同様に、ユーザは必要に応じてマイルームにもページを追加し、モジュールを配置することができる。たとえば、やるべき仕事 (ToDo) のリストや日々のスケジュールを管理したり、キャビネットを設置することで、論文のバックアップを管理したりできる。



図 9: ToDo とスケジュールを組み合わせてマイルームに配置した様子

これらのモジュールは国内主要 3 キャリアの携帯電話からアクセス可能である。出張先から携帯電話でスケジュールを確認することも可能であり、ユビキタス時代の研究者のツールとして進化することが期待できる。

(3) コミュニティ

研究者の共同研究支援を行う仕組みが「コミュニティ」である。

IDを持つ研究者は Researchmap 内に研究コミュニティを自由に作成し、他の研究者を活動に誘うことができる。まず、コミュニティの名前、目的、概要などを書き、申請を行う。次にコミュニティの公開レベルを設定する。これにより、すべての研究者に開かれたコミュニティとしても、クローズドなコミュニティとしても、あるいは招待制のコミュニティとしても設定することができる。

コミュニティには、第一次新領域融合研究プロジェクトで研究開発した融合研究を促進するための様々な機能（メーリングリスト配信機能、電子掲示板、共有キャビネット、共有カレンダー、スケジューラー、ToDo リスト、汎用データベースなど）がモジュールとして盛り込まれている。これらを自由に選択することで、その研究コミュニティに必要な情報の共有を即座に実現することができる。

共同研究者がまだ Researchmap の ID を取得していない場合には、「招待」機能を使って Researchmap への参加を促すことができる。招待できる研究者の数には制限はないので、自分が望むだけの研究者を Researchmap に誘うことができる。

コミュニティの活動は参加者のみに公開され、研究者以外には公開されることはない。また、コミュニティの管理者はサイト全体の管理者の許可なしに、コミュニティ参加者の権限を変更することができる。これにより、コミュニティ内のトラブルを管理者の手を借りずに、コミュニティ内で迅速かつ適切に処理することができる。

現在、Researchmap では 100 以上の研究コミュニティが構築され、共同研究や共同論文執筆等に利用されている。その中には、包括脳ネットワーク（科研費特定領域研究）、情報処理学会 WLP などのバーチャルコミュニティがある。

(4) ポータル

Researchmap 上に参加している膨大な数の研究者が発信する情報を、コンピュータが自動的に整理し、可視化する必要がある。それを担うのが、「ポータル」に設置された各種機能の役割である。研究者の氏名、所属のほか、研究キーワードから検索し、重要度順に表示する研究者検索のほか、アクティブな研究コミュニティを表示する「ホットコミュニティ」、全国で行われている研究イベントを集約表示する「研究イベント」などを提供している。このほかにも、研究者の論文や過去の研究活動から研究の近さを計算して表示する「おとなりの研究者」などの機能を提供している。

平成 22 年度末現在、Researchmap には 4 千人を超える研究者が参加している。平成 23 年度に ReaD と連携を開始することにより、20 万人を超える研究者が参加する世界有数の研究者サービスに成長する見込みである。

[5] 研究成果物

① 知見・成果物・知的財産権等

1. 「連想×書棚」システムの構築

連想の基点となる文脈を保持する「連想×書棚」システムを構築し、異分野研究者間の創造的対話における有効性を確認した。新 WebcatPlus にて公開（2010.6, <http://webcatplus.nii.ac.jp/>）。Researchmap 等の外部サービスから利用可能にするプラグパーツを作成した。

2. 人物情報を軸とする高信頼情報の収集・整理

ウィキペディアより人物項目 14 万人を収集し、人物ごとに関連する著作物を連想検索により自動整理した。新 WebcatPlus にて公開（2010.6）。

3. 連想計算による異種情報源の相互分類
WebcatPlus 書誌情報とウィキペディアを対象に、書籍群と関連するウィキペディア項目の収集、ウィキペディア項目語彙による特徴語抽出、ウィキペディアの項目間構造を用いた自動分類の実験を行った。新 WebcatPlus にて公開 (2010.6)。
4. 国立国会図書館サーチ (試行版) への連想キーワード機能提供
連想検索エンジン GETAssoc を利用して実装し、国立国会図書館サーチ (試行版) にて公開 (2010.8, <http://iss.ndl.go.jp/>)。
5. 美術館等の情報発信システム構築
徳川美術館、東京文化財研究所、慶応大学斯道文庫、小布施町における文化財情報の電子化に協力して、常設展や企画展での情報発信を分担した。小布施町と神保町においては、古地図と現代のイラスト地図を重ねて街歩きを楽しめる iPhone アプリを開発公開した。
6. 分野横断型研究会における知識共有実験開始
「人類の移動誌」研究会を対象に、研究データや研究会資料の電子化に着手した (<http://idoushi.jp/>)。研究会ホームページは NetCommons を活用して作成した。
7. Researchmap の開発とリリース

② 成果発表等

<論文発表>

[学術論文]

1. Sugawara, T. Ishijima, and N.H. Arai, Method of Training K-12 Students Skills to Convey Information to Others in the Web2.0 Society, S. Proceedings of the 13th IASTED conference on Computers and Advanced Technology in Education (CATE2010), p 110- p 117, 2010 年 8 月
2. K. Kawamoto and N.H. Arai, Developing the Logical Thinking Ability by Introducing Learners' Mutual Evaluation on an Electronic Bulletin, Proceedings of the 13th IASTED conference on Computers and Advanced Technology in Education, p69- p 76, 2010 年 8 月

[データベース]

[著書等]

1. 『コンピュータが仕事を奪う』, 新井紀子執筆, 日本経済新聞出版社, 2010 年

[解説・総説]

1. 橋田浩一, 武田英明, 長尾確, 小橋康章: ソーシャル e サイエンスと総合学術オントロジー, 人工知能学会誌, Vol. 25, No. 4, pp. 581-588 (2010-07-01).
2. 武田英明: 日本における Linked Data の現状と普及に向けた課題, 情報処理, Vol. 53, No. 3, pp. 326-333 (2011).
3. 高野明彦, 現代思想(vol.39, no.1 p.72-85) 特集=Google の思想 グーグル的なものとどうつき合うか, 2011(H23).1.1

[その他]

1. L. Xu, H. Takeda, M. Hamasaki and H. Wu: Typing Software Articles with Wikipedia Category Structure, Technical Report NII-2010-002E, National Institute of Informatics, Tokyo (2010).

2. 新井紀子, 野崎昭弘, 亀井哲治郎, 数学文化公開座談会より 数学のどこでつまづくか, 数学文化 (14) 30-50 2010 年 7 月
3. 新井紀子, いやでも役に立つ数学 中等教育資料 (889) 2-3 2010 年 6 月

<会議発表等>

[招待講演・国際]

1. H. Takeda: The key for the Information Society - the Information Cycle: Create, Publish and Share, in The First International Workshop on Semantic Web, Mobile Web, and Social Networks, KAIST, Daejeon, Korea (2010).
2. H. Takeda: Connecting Museums with Linked Data, in When Culture Encounters Internet Conference, Taipei (2010).
3. 高野明彦, 「東アジア情報と中国語新聞電子化のための図書館協力に関する国際シンポジウム」, 「From Search to Association -- Web service to bridge the isolated silos of knowledge (招待講演)」, 中国・長沙, 2010 年 10 月 25 日
4. 高野明彦, 「Japan-Vietnam Workshop on Software Engineering 2010」, 「From Search to Association -- How to extract inspiration from data (招待講演)」, ベトナム国立大学, 2010 年 12 月 9 日
5. 新井紀子「Researchmap Project (招待講演)」, NSF 50th anniversary symposium 2010 年 10 月 National Science Foundation

[招待講演・国内]

1. 武田英明: Linked Data アプローチによる芸術情報統合の試み, シンポジウム「アーカイブから紡ぎ出される知」東京藝術大学 芸術情報センター (2010).
2. 武田英明: 日本における Linked Data の課題とその解決への試み, JEITA 知識情報処理技術に関するシンポジウム「リンクするデータ, リンクするサービス」電子情報技術産業協会 (2011).
3. 高野明彦, 「連想検索: 知の蔵を繋ぐための情報サービス」第 30 回ソフトウェア・シンポジウム (ソフトウェア技術者協会主催), 2010 年 6 月 9 日,
4. 丸川雄三, デジタルアーカイブがひらく未来 (基調講演), 「デジタルアーカイブシンポジウム」(小布施町立図書館主催), 「小布施町立図書館, 2010 年 7 月 19 日,
5. 高野明彦, 「第 9 回情報科学技術フォーラム」(社団法人 電子情報通信学会, 一般社団法人情報処理学会共催), 「『連想×書棚』で知識の扉を開く」, 九州大学, 2010 年 9 月 8 日,
6. 高野明彦, 「日本電子出版協会 平成 23 年度通常総会セミナー」(日本電子出版協会主催), 「読書環境の未来について」, 日本教育会館, 2010 年 2 月 24 日,
7. 新井紀子, 「数学は言葉」長野県立屋代高校 SSH 基調講演 2010 年 5 月
8. 新井紀子, 「ネットの上に学びの場を創る」北海道大学高等教育機能開発総合センター科学技術コミュニケーション教育研究部特別講演 2010 年 5 月
9. 新井紀子, 「いやでも役に立つ数学」, 札幌三省堂サイエンスカフェ 2010 年 5 月
10. 新井紀子, 「デジタル時代の教育を考えるデジタル時代の教育を考える」 2010 年 9 月 読売新聞社, 文字・活字文化推進機構
11. 新井紀子, 「科学の言葉としての数学」サイエンスカフェ 2010 年 10 月 日本学術会議
12. 新井紀子, 「次世代へのメッセージ」ノーベル賞フォーラム 2010 年 10 月 読売新聞社
13. 新井紀子, 「21 世紀の学びを支える情報化とは～校務・研修のクラウド化について」全国教育研究所連盟教育課題研究協議会 2010 年 11 月 全国教育研究所連盟

14. 新井紀子, 「いま, 本当の情報化について話そう」
教育フォーラム 2011 2011年1月 東京学芸大学・3市連携 IT 活用コンソーシアム
15. 新井紀子, 「科学の言葉としての数学」技術同友会例会 2011年3月

[一般講演・国際]

1. Kiyoko Uchiyama, Hidetsugu Nanba, Akiko Aizawa and Takeshi Sagara, OSUSUME: Cross-lingual recommender system for research papers, Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation, pp.39-42, Stanford, 2011.
2. Makoto Miwa, Yusuke Miyao, Rune Sætre and Jun'ichi Tsujii, Entity-Focused Sentence Simplification for Relation Extraction. In the Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). pp. 788-796. 2010.
3. Yusuke Miyao, Alastair Butler, Kei Yoshimoto and Jun'ichi Tsujii, A Modular Architecture for the Wide-Coverage Translation of Natural Language Texts into Predicate Logic Formulas. In the Proceedings of PACLIC 24. pp. 481-488. 2010.
4. Atsuhiko Takasu, Cross-lingual keyword recommendation using latent topics, International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2010), pp.52-56, 2010.
5. Pakapon Tangphoklang, Saranya Maneeroj, Atsuhiko Takasu, A Novel Weighting Scheme for a Multi-Criteria Rating Recommender System, IADIS International Conference on Information Systems (IS2011), pp. 21-29, 2011.(Best Applied Research Paper)

[一般講演・国内]

1. 内山清子, 専門用語の分野基礎性に関する一考察, 情報処理学会自然言語処理研究会, 2010-NL-199(15), pp1-6, 広島市立大学, 2010年11月19日.
2. 内山清子, 専門分野における用語の分野基礎性に関する研究, 言語処理学会第17回年次大会, pp.1033-1036, 2011年3月10日.
3. 相澤彰子, 情報検索における圧縮距離の適用に関する考察, 情報処理学会自然言語処理研究会, 2010-NL-199(8), pp1-10, 広島市立大学, 2010年11月18日.
4. 丹英之, 大向一輝, 武田英明: 日本語リポジトリ「ことばぶ」の構築, 人工知能学会全国大会(第24回)論文集, No. 2B1-2, 長崎 (2010), 人工知能学会.
5. 嘉村哲郎, 加藤文彦, 大向一輝, 武田英明, 高橋徹, 上田洋: Linked Open Dataによる多様なミュージアム情報の統合, 人文科学とコンピュータシンポジウム論文集, 情報処理学会シンポジウムシリーズ, 第2010巻, pp. 77-84 (2010).
6. 深見嘉明, 小林巖生, 嘉村哲郎, 加藤文彦, 大向一輝, 武田英明, 高橋徹, 上田洋: Linked Data とコミュニティが拓くオープンガバメント, 情報社会学会 第3回知識共有コミュニティワークショップ (2010).
7. 深見嘉明, 小林巖生, 嘉村哲郎, 加藤文彦, 大向一輝, 武田英明, 高橋徹, 上田洋: Linked Open Dataによるボトムアップ型オープンガバメントの試み, デジタルドキュメント研究会 第79回研究会, 情報処理学会 (2011).

[ポスター]

1. 福田悟志, 難波英嗣, 竹澤寿幸. “技術文書からの動向情報の抽出と可視化” 言語処理学会 第17回年次

大会, pp.276-279, 豊橋技術科学大学, 2011年3月8日.

2. H. Wu, T. W. Ling, L. Xu, H. Takeda and M. Hamasaki: An Adaptive Ontology-based Approach to Identify Correlation between Publications, in *Poster Proceedings of 20th International World Wide Web Conference (WWW2011)* (2011).

<受賞>

1. Pakapon Tangphoklang, Saranya Maneeroj, Atsuhiko Takasu, A Novel Weighting Scheme for a Multi-Criteria Rating Recommender System, IADIS International Conference on Information Systems (IS2011), pp. 21-29, 2011.(Best Applied Research Paper)
2. 新井紀子, 文部科学省科学技術分野の文部科学大臣表彰(理解増進部門), 2010年4月
3. N. Arai and R. Masukawa, Researchmap Opening the Door to the World of Science2.0, Proceedings of the 13th IASTED conference on Computers and Advanced Technology in Education (CATE2010), selected among three best papers at CATE2010, 2010年8月

③ その他の成果発表

<新聞報道など>

1. 高野明彦, 2010(H22).4.30, 日刊工業新聞(朝刊 10面) 古書店 滅亡か, 復権か- デジタル化の道探る
2. 高野明彦, 2010(H22).4.30, 日刊工業新聞(朝刊 10面) 流通・著作権の統一規格整備へ 文科・経産など3省
3. 高野明彦, 2010(H22).5.31, 日本経済新聞(夕刊 16面) 新しい古書街へ進化する神保町
4. 高野明彦, 2010(H22).6.4, サンケイスポーツ(18面) 古きを訪ねて本との新しい出逢い
5. 高野明彦, 2010(H22).6.18, NII press (国立情報学研究所 報道発表) 新 Webcat Plus : 確かな知識の基点サービス
6. 高野明彦, 2010(H22).6.19, 朝日新聞(夕刊 14面) 2000万冊 ネットで探せます 検索システム無料公開 一部の購入可能
7. 高野明彦, 2010(H22).6.19, asahi.com 書籍2千万冊の検索システム 無料公開 一部は購入OK
8. 高野明彦, 2010(H22).6.21, INTERNET Watch 国立情報学研究所, 約1900万冊の書誌情報検索サービス「Webcat Plus」 http://internet.watch.impress.co.jp/docs/news/20100621_375942.htm
9. 高野明彦, 2010(H22).7.8, 信濃毎日新聞(朝刊 33面) 小布施の今昔持ち歩こう 現代・江戸時代の地図から情報検索 アイフォン・アイパッド向け 町立図書館ソフト開発
10. 高野明彦, 2010(H22).7.9 科学新聞(5面) IT・情報通信動静
11. 高野明彦, 2010(H22).7.21 信濃毎日新聞(朝刊 13面) 文化資料創造的利用探る デジタルアーカイブ意義と可能性は 小布施でシンポジウム 「小布施人百選」事業紹介 100年後に知恵届ける
12. 高野明彦, 2010(H22).8.18 INTERNET Watch 「国会図書館サーチ」開発版が公開, 全国図書館の蔵書や資料などを検索可能
http://internet.watch.impress.co.jp/docs/news/20100818_387639.html
13. 高野明彦, 2010(H22).9.2 朝日新聞(朝刊 13面) (全面広告) MARUZEN&ジュンク堂書店 ユニークな「連想検索」システムを初導入
14. 高野明彦, 2010(H22).10.8, 東京サイト(13:55~13:59) 放送 本の楽しみが広がる街
15. 高野明彦, 2010(H22).10.10, 神保町公式ガイド vol.1(p.34-35) 本と街の案内所 “もう道に迷うなかれ!”
16. 高野明彦, 2010(H22).10.25, REDNET.CH (中国メディア) "原文:【首发】图书馆馆长会议开幕 200余

专家聚焦报刊文献数字化

訳：【開幕】オープニングセッションには、200人以上の専門家が出席し、新聞のデジタル化について話し合われた。"

http://hn.rednet.cn/c/2010/10/25/2095891_1.htm

17. 高野明彦, 2010(H22).10.30, ウィークエンド中部 NHK(7:30~8:00 放送) リポート 深まる秋, 本を楽しもう!
18. 高野明彦, 2010(H22).11.01, てんとう虫(vol.42, no.11 (no.553) p,24-24) "大型書店のコラボで誕生した質量ともに充実の""知の回廊"", MARUZEN&ジュンク堂書店 渋谷店"
19. 高野明彦, 2010(H22).11.02, てんとう虫(vol.42, no.11 (no.553) p,28-28) ネットを使って紙の本と出会う
20. 高野明彦, 2010(H22).11.03, てんとう虫(vol.42, no.11 (no.553) p,30-31) 神保町を歩く 楽しきかな, 本の街
21. 高野明彦, 2010(H22).11.05, NII press (国立情報学研究所 報道発表) 黒田清輝関係写真アーカイブ「写真で見る黒田清輝の日常」を一般公開
22. 高野明彦, 2010(H22).11.06, 信濃毎日新聞(日刊 25 面) オピニオン デジタル時代 あなたの読書は? シンポジウムのパネリストから
23. 高野明彦, 2010(H22).11.09, NHK ラジオ 第1 私も一言! 夕方ニュース 17:00~18:50 放送 ニュースここの一番「グーグルの検索シェア 9割に物言い ~あなたは今のネット検索に満足していますか?」
24. 高野明彦, 2010(H22).11.12, NII press (国立情報学研究所 報道発表)
<http://www.nii.ac.jp/news/2010/1112/>
タッチパネルと高精細画像で国宝を鑑賞できる 「Powers of Information 徳川美術館」を一般公開
25. 高野明彦, 2010(H22).11.13, 徳川美術館お知らせ
http://www.tokugawa-art-museum.jp/news/detail_146_1.html
タッチパネルと高精細画像で国宝を鑑賞できる 「Powers of Information 徳川美術館」を一般公開
26. 高野明彦, 2010(H22).11.22, 文教速報 no.7518 p.13-13 タッチパネルと高精細画像で国宝を鑑賞 NII 徳川美術館と共同でシステム開発
27. 高野明彦, 2010(H22).11.25, 読売新聞 朝刊(28 面) 65 インチ画面で作品鑑賞 徳川美術館 12 分野の 850 点
28. 高野明彦, 2010(H22).11.30, NII press(国立情報学研究所 報道発表)
<http://www.nii.ac.jp/news/2010/1130/>
タッチパネルと高精細画像で書誌学の世界にふれる 「Powers of Information 斯道文庫」を一般公開
29. 高野明彦, 2010(H22).12.03, NII press (国立情報学研究所 報道発表)
<http://www.nii.ac.jp/news/2010/1203/>
国宝や重要文化財を網羅, 文化遺産を多様な切り口で検索できる 「文化遺産データベース」を一般公開
30. 高野明彦, 2010(H22).12.03, NDL Current Awareness Portal
<http://current.ndl.go.jp/node/17220>
「文化遺産オンライン」で「文化遺産データベース」が一般公開
31. 高野明彦, 2010(H22).12.04, 信濃毎日新聞(日刊 25 面) オピニオン デジタル時代 あなたの読書は? 「Wa の会」小布施の討論会から(上)
32. 高野明彦, 2010(H22).12.05, 朝日新聞 16 面 (日曜日)望む 変化のときに 情報で人と人, 地域と地域を結ぶ まちとしょテラソ館長 花井裕一郎さん(48) 図書館の可能性広げる
33. 高野明彦, 2010(H22).12.11, 信濃毎日新聞(日刊 26 面) オピニオン デジタル時代 あなたの読書は?

「Wa の会」小布施の討論会から(下)

34. 高野明彦, 2010(H22).12.13, 日本情報産業新聞(朝刊 2 面) データベース公開 文化遺産の検索可能に文化庁 NII
35. 高野明彦, 2011(H23).1.17, 西日本新聞(朝刊 10 面) 書物の行く末 5 図書館 知的生産の未来を担う
36. 高野明彦, 2011(H23).1.22, 週刊 東洋経済(no.6304, p.54-56) 日本でただ 1 冊の希少本を探せ中古本最前線 古書店は各店が個性的 時間をかけて回るべし
37. 高野明彦, 2011(H23).2.13, 日本経済新聞(電子版ニュース) ウィキペディア誕生 10 年 信頼性向上へ悪戦苦闘 専門家が精査, 学生・主婦らが勉強会
38. 高野明彦, 2011(H23).3.10, プリバリ印(vol.25 p.38-39) クロスメディア時代の出版印刷 連想検索がもたらすコンテンツとの出会い
39. 高野明彦, 2011(H23).3.31, ノジュール(2011 年 4 月号 no.54 p.18) 本と街の案内所 先端技術で目当ての本探し
40. 新井紀子, コンピューターに勝つには, 2011/03/28 日経ビジネス P.78
41. 新井紀子, Review コンピュータが仕事を奪う, 2011/02/19 週刊東洋経済 2011.2.19 p.120
42. 新井紀子, 教員研修や情報教育 考える研究会, 開幕=岩手 2010/11/05 東京読売新聞 朝刊 31 ページ
43. 新井紀子, ノーベル賞受賞者を囲むフォーラム 「次世代へのメッセージ」 発見生む想像力 海外へ飛び出せ 2010/10/26 読売新聞 朝刊 20, 21 面
44. 新井紀子, ノーベル賞受賞者フォーラム 熊本で「雑学の楽しみ身につけよう」2010/10/13 西部読売新聞 朝刊 1 ページ
45. 新井紀子, ノーベル賞討論会に 560 人 2010/10/13 東京読売新聞 朝刊 37 ページ
46. 新井紀子, シンポジウム「デジタル時代の教育を考える」 論争 デジタル教科書「学ぶ力」低下に懸念 2010/09/19 読売新聞 朝刊 22, 23 面
47. 新井紀子, 数学者と学習交流 熊谷高校で成果を発表 2010/09/15 埼玉新聞 朝刊 13 面
48. 新井紀子, 数学をテーマに生徒が研究発表 熊谷高 2010/09/15 朝日新聞 朝刊 34 面
49. 新井紀子, 【街ふれあい】「計算」を探求 県立熊谷高校で数学探求教室「計算とは何か」 学習成果発表会 2010/09/12 読売新聞 朝刊 33 面
50. 新井紀子, 鈴木文科副大臣, 小学校教科書のデジタル化無いと表明 2010/09/08 文化通信速報版 5 ページ
51. 新井紀子, 9 月 3 日, 東京で デジタル教科書考えるシンポ 2010/07/27 読売新聞 朝刊 38 面
52. 新井紀子, 2010 年度日本 OSS 貢献者賞, IPA が候補者の推薦募集を開始 2010/07/16 IT Pro オンライン
53. 新井紀子, ナゾ謎かがく=数学なぜギリシャで発展? 論理力は都市国家の”武器” 2010/06/27 日本経済新聞 朝刊 17 面
54. 新井紀子, 平成 22 年度科学技術分野の文部科学大臣表彰 2010/04/06 東京読売新聞 朝刊 28 ページ