

プロジェクト名： 分野横断型融合研究のための情報空間・情報基盤の構築 [横断型融合]

プロジェクトディレクター： 東倉洋一

1. 研究目標

本プロジェクトは、生命システム、地球システムを具体的な対象として、大規模で複雑な巨大システムの解明に向けて、「生命」、「地球」、「統計数理」、「情報」の4分野横断型の融合研究を加速的に推進するための情報空間・情報基盤を確立することを目標とする。

2. 研究概要

従来、統計数理や情報と他研究分野との融合は、統計数理や情報におけるさまざまな理論、手法、技術などを、他研究分野に取り入れることによって、研究の推進や新しい展開を意図したものが一般的であった。しかし、統計数理や情報を研究の手段として利活用するだけの研究では必ずしも十分な分野融合が期待できず、その成果も限られた範囲に留まることが少なくない。本プロジェクトでは、異分野を十分に融合させ、融合研究の深化と効果的な推進を加速させることを重視した新しい形の融合研究を追究する。具体的には、融合研究によって、従来の手法では見ることができなかった生命や地球の新しい姿を明らかにするための情報空間・情報基盤の確立、および、統計数理や情報の理論、手法、技術における新しい展開の2つの視点の両立を図ることを重視する。

研究姿勢として、「現場を知り」「現場に学び」「現場を分かり」「現場を創る」ことをコンセプトとして研究を進める。「現場」とは、生命・地球システムに代表される巨大システムである。この巨大システムへの挑戦には、システムを構成する大規模で異分野・異種の情報が複雑に絡み合った情報空間の構造を解明し、ここに含まれる情報を有効かつ効率よく活用するための手法が必要となる。本プロジェクトでは、このような大規模・異種情報に対して、「連想」と「リンケージ」という切り口で取り組む。また、生命・地球システムの研究では、生命情報や地球情報に関する多様で大規模なデータが必要であり、量と質に関して十分に要求条件を満たす統合データベースや統合的情報基盤の構築と活用手法の研究開発を行う。さらに、「生命」、「地球」、「統計数理」、「情報」の4分野にわたる横断型の共同研究や研究協同を進めるためのコラボレーションやコミュニティ形成を促進する情報共有基盤が不可欠であり、これをバーチャルラボとして構築することを目指す。

すなわち、生命や地球の研究に新たな発見をもたらすとともに、統計数理や情報の研究における理論、手法、技術の壁や限界を突破すること、さらには未踏研究領域の開拓、未来価値の創成が、本プロジェクトで推進する融合研究の究極的な狙いとなる。

上記のプロジェクトを推進するため、具体的には、3つのサブプロジェクトを実施する。すなわち、サブテーマ1：大規模・異種情報の収集・解析・結合・分類の手法および知識基盤の構築、サブテーマ2：地球・生命などの巨大システム解明のための統合的情報基盤の形成と活用手法の

確立，サブテーマ3：コラボレーションとコミュニティ形成のための情報共有基盤とバーチャルラボの構築である。さらに，サブテーマ1は，「知識発見のツール構築」，サブテーマ2は，サブテーマ1で構築した知識発見ツールを用いた「知識発見・原理解明」，サブテーマ3は，サブテーマ1と2を支える「知識共有基盤整備」と位置付け，「融合知」に向けたサブテーマ間の連携・融合を図る。

3. 年度計画

テーマ	16年度 予備研究	17年度 プロジェクト初年度	18年度	19年度 中間評価	20年度	21年度
サブテーマ1	←				→	→
サブテーマ2	←				→	→
サブテーマ3	←				→	→

平成16年度（予備研究）

サブテーマ1に関しては，異種情報の結合・分類手法の基礎となる連想検索の研究開発を進め，文書情報の類似性を高速に計算する汎用連想計算エンジンGETA(Generic Engine for Transposable Association)の開発と，そのオープンソース形態での配布を行ってきた。また，GETAを利用した図書情報検索サービスWebcat Plusによる，900万冊の本の連想検索の提供をはじめとした多様な情報サービスへの適用によって，要素技術としての充実を図ってきた。

サブテーマ2に関しては，ライフサイエンスメタデータベースシステム構築の研究を推進するための枠組みとなるバイオポータルの研究開発を国立情報学研究所と国立遺伝学研究所が共同で推進し，これを公開した。最新の研究成果を「科学者がわかりやすい日本語で語る」ことを基本として，専門用語の変換翻訳，用語辞書，解説などの整備を行い，研究成果の背景にある生命の原理に触れるための道具として提供した。

サブテーマ3に関しては，情報共有基盤の基礎となるグループウェア「NetCommons」を構築し，全国約70団体で運用実験を行ってきた。また，大学院教育を主とする高等教育向けe-LearningシステムWebELS(Web-based e-Learning System)の出発点となるWebLSの研究開発を行い，試用を行ってきた。

平成17年度（プロジェクト開始）

プロジェクト開始にあたって，プロジェクトの目標と方向性の明確化，サブテーマにおける研

究項目と研究計画の具体化を行い、3つのサブテーマ毎の要素技術に関する研究開発を、サブテーマ間の連携と生命・地球システム解明を目指した情報基盤構築という目標を共有して推進する。プロジェクト運営において、特に、機動性と柔軟性を重視した予算配分・経費投資および人材雇用・補強を重視する。

サブテーマ1では、分野を横断して存在する異種データの情報内容の類似性に基づく連想的な結合・統合と自動分類などの手法の検討とシステム実装、および、実証実験に用いる各種情報を収集・整備する。これらの情報を使って、異種情報源の関連づけとその提示が可能な利用環境のプロトタイプの開発を行う。また、リンケージ情報を機械的かつ大規模に収集するための機械学習・データマイニング等の要素技術の研究開発を行う。

サブテーマ2においては、ライフ系統合データベースの核となるメタデータベースシステム構築のための基本設計、および、必要なハードウェアやソフトウェアの整備を行い、国内データベース充実に力点を置いたシステム構築を開始する。また、国立極地研究所が収集してきた極地生物資料のうち、植物関係の資料を対象とし、従来から付記されている情報に新しい情報を加えた統合データベース構築のための実験、データベースの基本設計、および、必要な設備等を整えるとともに、貴重資料に対して適切な手法の開発を開始する。さらに、多種多様な地球環境データの統合について検討し、統合に関する方向性と基本的な考え方を明確にするとともに、いくつかの統合形態について、Web上のデータを使った実験的な検討を行う。

サブテーマ3においては、バーチャルラボ構築支援システム NetCommons を、情報・システム研究機構が有する多様なWebアプリケーションと連携させ、大規模データベースを分散環境で活用しながら効率的で実用的な共同研究システムに発展させることを主たる目標とする。さらに、大学院教育を主とする高等教育向けe-LearningシステムWebELS (Web-based e-Learning System) の分野融合研究およびその成果の教育展開を目指す。

平成18年度

サブテーマ1においては、連想的に結合された異種データの有用性を示すため、環境問題を例題に取り、論文、専門辞典、教科書、新聞記事、書籍、特許など多様な情報を、テーマ別に自動分類する方式について研究する。また、研究者間のコミュニケーションがますます困難になりつつある生命科学分野で、困難さの原因が固有名称の多用や遺伝子の機能構造に関する自然言語表現にあることに注目して、それらを自動的にアイコン化して分野間のギャップを埋めるジーンアイコン (遺伝子象形文字) プロジェクトを推進する。さらに、リンケージ情報を機械的かつ大規模に収集するための機械学習・マイニングの要素技術を研究するとともに、国立情報学研究所の科学研究費補助金データベースを利用した研究者基礎データの構築、および、日米の引用索引データベースを利用した学術構造の分析について調査研究を進める。

サブテーマ2では、ライフサイエンスメタデータベースシステムへのデータ登録作業を続行するとともに、登録済みデータベースについて機械処理およびエキスパートを使って評価を実施する。また、極限環境生物統合データベース構築に必要なソフトウェアの改良とデータ収集を進め、

Web 公開用のプロトタイプを作成する。極地研収集資料の中で、特に2種類の汎地球種のサンプリングを先行実施し、約1000検体の回収を目指す。また、実験動物(マウス)を用いた皮下脂肪、内臓脂肪定量試作解析システムの評価を行い、必要な研究開発項目の追加と計測解析手法の修正・改良を実施する。

さらに地球情報に関しては、地球科学の各分野(各グループ)で観測したデータを多目的に利用するための地球環境ポータル「鉛直地球(Vertical Earth)」の構築を継続する。地球科学データベース「南極GIS」に関して、地理情報システムに基礎的な地図データ等を蓄積・利用可能なレベルを達成するとともに、データベース統合に関して、「鉛直地球(Vertical Earth)」のプロトタイプを公開可能な状態にまで整備する。

サブテーマ3では、ユーザ間の契約に基づく情報共有のモデルの出発点として、情報共有システム基盤間の認証モデルを構築し、プロトタイプを実装する。また、ユーザが興味を持ったWeb情報の「擬似記憶」モデルの研究開発を行い、NetCommons上に実装する。これを共同研究機関等に提供し、そのユーザビリティ等についてのフィードバックを受ける。次に、Web2.0の動向を見据え、コアプログラムの再検討を行うとともに、ブラウザを意識させないユーザインタフェースとより迅速な情報伝達を目指す。

WebELSについては、総研大のe-Learningプラットフォームとしての研究開発およびサービスの提供、清華大学、チュラロンコン大学の教育利用とタイ企業との開発・サービス協力の他、国際連携による利用を促進する。また、国内大学の利用希望者へのシステム提供・利用の方法や利用者が研究開発に参加出来るフレームワークを検討する。

平成19年度 (中間評価)

サブテーマ1では、環境関連情報と生命科学分野の研究情報を例にとり、異種情報源から情報内容の類似性に基づいて関連情報を収集し、それらを概観しやすい形で提示する情報システムを試作する。専門辞典などを軸に関連情報を動的に整理して提示する。ジーンアイコンを活用して、文献の深い理解に必要な遺伝子等の情報について、最新の関連データの内容を略図表示するシステムも試作する。

また、平成19年度に試作した研究者情報サーバを中核として、書誌データベースやWebなど外部の情報源との情報統合について検討を進める。特に、別途開発した書誌同定サーバと連携させ、論文の著者IDを自動認識し統合するための手法の確立を目指す。また、研究課題の代表者と分担者の関係に基づく研究者ネットワークを構成し、研究者コミュニティの抽出や類型化等のネットワーク分析を行う。名簿マッチングに代表されるように人を中心としたコミュニティ相関分析について理論面での検討を行うとともに、引き続き学術構造分析についての研究を進める。

サブテーマ2では、ライフサイエンスメタデータベースシステム構築では、ライフ系ファクトDBについてのデータ収集作業を継続すると共に、専門知識を持つ人材による独自のアノテーション付加作業も継続し、最終的にはサブテーマ1が開発したイメージシステムへの適用を試みる。また、ライフサイエンス統合DBに関するJST事業とも連携し、データの相互利用形態についての

検討を実施する。地球ポータルと南極 GIS の構築では、「鉛直地球ポータル」をリニューアル公開するとともに、鉛直層構造のなかの一層に関して、データの閲覧と検索が可能な状態にする。

サブテーマ 3 では、複数の NetCommons 間で認証を行い、コンテンツの移転・コピー・エイリアスの作成などを契約に基づいて安全かつ簡便に行うモデルを構築・実装するとともに、NetCommons を個人のバーチャルデスクトップ・ファイルサーバとして利用するための研究開発を行い、これを NetCommons ver2.0β として提供する。NetCommons ver2.0 の大規模実証実験を行い、特に、研究者が端末や場所に束縛されずに常にシームレスに研究を継続できる環境を提供する。学校 Web サイトの ASP サービスの実証実験の成果をまとめ、コンソーシアムに知見を蓄積し、国内外に配信していくとともに、多言語化の充実を図る。特に清華大学との MOU を通じ、中国語化を行い、中国語版 NetCommons の拠点を清華大学に移転することを目指す。全国の小中高校の学校 Web サイト構築基盤部門におけるトップシェアを目指す。

また、総研大のテラーメード教育推進計画との連携をより強化し、基盤 e-Learning プラットフォームとしての要求に応えられるよう機能の拡張・改良やサービスの質的向上を図るとともに、MOU 提携校である清華大学、チュラロンコン大学、ダッカ大学等との国際連携を深める。これらを通して、実利用と評価に基づく仕様の見直しと柔軟な研究開発を推進するとともに、本研究が、総研大において自立推進可能な基盤の構築を目指す。また、企業や大学における教育研修ツールとしての実用性評価を行うほか、研究・教育の国際化を促進する時代の要望に添って国内大学等への利用の展開も図る。

平成 20 年度

サブテーマ 1 では、異種情報の結合・分類手法の研究においては、専門性の極端に異なる情報源の間での連想計算について追求する。これには、専門辞典における用語の説明文を手がかりに、専門性の高い用語の内容を一般的な言葉で表す方法を検討する。また、専門性の極端に異なる情報源の間では用語の違いによる類似性を見落としに関して検討する。さらに、用語集合が極端に異なる例の一つとして、日本語版と英語版のウィキペディアを取り上げ、その間の連想計算の精度向上を目指す。

大規模リンクエッジ情報の研究では、平成 19 年度に統計分野を対象に試作した研究者同定ツールを、他分野に応用させる方策を検討し、分野に関わらず研究者を同定する枠組みを確立する。リンクエッジエンジンとあわせて、研究者およびその研究成果としての書籍・論文間の情報リンクエッジの全体像を確立する。これと並行して、研究者ネットワークと研究者発信コンテンツの内容分析を組み合わせた研究者推薦システムのプロトタイプを試作する。さらに、これまで培った情報リンクエッジ手法を Web 上の情報源に適用する方策を検討する。また、引き続き分野毎の差異や経年変化についての分析を行うとともに、得られた全体像を俯瞰して分析可能とするツールの試作を検討する。

サブテーマ 2 では、ライフサイエンスメタデータベースシステム構築研究では、新たに創薬利用に向けた検索システムの開発を行う予定である。年度内にプロトタイプを完成、次年度に実用

システムとしてバイオポータルサイトからの公開を目指す。

極限環境生物統合データベースの構築研究では、これまで DNA 解析実験の場であった理化学研究所ゲノム科学総合研究センターが19年度末で廃止されることに伴い、20年度以降は国立遺伝学研究所との連携を中心にゲノム解析を進める予定である。コケ類のゲノムについては平成19年に代表種としてヒメツリガネゴケゲノムの解読にともない、コケ類を対象としたゲノム研究を、試料の保存と採取については国立極地研究所、データ生産については国立遺伝研究所、大量情報処理については国立情報学研究所という共同研究体制で推進する。また、平成20年度からは東京工業大学と連携し、新たな極限環境生物としてシーラカンスのゲノム解析を取り上げる。

地球環境データ統合データベース研究では、「鉛直地球 (Vertical Earth)」の機能を引き続き増強することにより、鉛直統合をコンセプトとしたデータ統合を進めていく。まず、過去の GPV データのアーカイブを用いて鉛直方向の台風データを蓄積したデータベースを構築する。これによって鉛直方向の構造の特徴を用いた台風の分類や検索を可能とする。構築した台風の鉛直構造データと、これまで蓄積してきた台風の水平構造データ (気象衛星画像) とを統合することによって、台風の立体構造を3次元CGとして可視化する。さらに、オントロジーを利用した地球科学情報の概念化にも取り組み、地球科学分野の独自のオントロジーを構築して Vertical Earth に取り入れる。南極 GIS については、サーバの保守・調整を継続して行い、GIS 上で基礎となる地図データの更新とその他の組み込み可能なデータを順次 GIS に組み込んでいく。システムについては、国土地理院等のデータとの調整を図り、一般公開用のサーバも実施する。

サブテーマ3では、Web2.0時代以降のワンストップシステムに関する研究開発を行う。現状のブラウザによる情報取得の形式は、主としてオンラインパブリッシングを想定したものであり、双方向あるいはマルチ方向であらゆる情報をやりとりすることを想定すると制限が非常に大きい。真にバーチャルラボシステムを志向するのであれば、X ウィンドウシステムに近い方法でポートレットを独立させて運用することが処理速度上もユーザインタフェースの観点からも望ましい。よって、ブラウザに代わる情報伝達ツールの検討やそのモデルの構築を行うとともに、その検討結果をブラウザの枠内で部分的に実現し、平成20年に公開予定の NetCommons2.0 に搭載する。

WebELS に関しては、総研大での本格利用に向けた実用性の向上と完成を目指す。具体的には、学生管理、正規科目管理を行うためのシステム管理者機能の強化、セキュリティ機能の強化、Internet 会議および遠隔講義機能の強化、ビデオファイル編集配信機能の強化、コンテンツ開発機能の強化、システムソフトの保守性の向上を図ると共に、GNU GPL 準拠によるオープンソフトサービス性の改善を図る。ユーザの新しい要望にも対応する。また、日本学術会議東アジア化学イニシャティブ分科会に設置されたグローバル複素大学コンソーシアム (GUC) 検討グループに求められる基盤ソフトとしての期待に応えられる国際展開を図る。さらに、東アジアの MOU 提携校との連携を推進し、UNESCO アジア本部との協力関係を強化すると共に、オーム社や NPO 法人日本教育振興協会との連携によって国内大学等への利用展開に努め、コンテンツビジネスや技術サポートビジネスを育成し、コンソーシアム構築を図る等の整備を行う。

平成21年度

サブテーマ1では、大規模な異種データ群に対して、論文（学術論文500万件のフルテキスト等）、教科書（大学講座シリーズ等）、新聞データ（600万記事）をはじめとする信頼できるデータベースをコア・データベースとする情報の、類似性に基づく連想的な結合・融合と自動分類を可能とする。

平成20年度で試作した推薦システムのプロトタイプを評価し、公開可能、限定的公開可能な情報を選別した上で試行運用を開始する。計算機により自動生成したデータは、コスト／スケール／網羅性の面では圧倒的に優位であるが、情報の誤りや欠落を多く含むことも事実である。品質にばらつきのある大量の情報が流通する今日において、同様の状況は様々な分野・局面において見られ、「少数・高品質」ではなく、「多数・低品質」なデータからの価値創出は、これからの情報社会における重要なポイントの1つである。情報学と統計学の融合領域を目指す本プロジェクトを通して、この問題に踏み込むためのアプローチを模索して行きたい。

サブテーマ2では、ライフサイエンス統合DB事業との連携・相互利用形態を検討・開始する。極限生物ゲノムについては、新世代並列型シーケンサを駆使することにより得られる大量のゲノム配列データを対象に、生物学、情報学の両側面から新たな切り口での解析を目指した研究を行う。また、鉛直地球ポータルでは、ウェブサービス活用による南極GISとのシステム統合のプロトタイプを作成する。統合したデータの変換機能、および、新世代地球ブラウザに対応した閲覧機能をポータルに実装する。

サブテーマ3では、NetCommonsに関して、これまでの実績を基盤に、本研究が、学校webサイトのSaaSサービスとして、継続的なサービスを実施するための事業化の形態を検討し、開始準備を行う。特に、平成20年度に新領域融合研究センターで開始予定の「研究者のためのSNS（researchmap.jp）に関する研究」と連動して、全国の研究者の研究情報を公開し、さらに共同研究の基盤を提供するようなSaaSサービスの提供を開始する。本サービスは長期間の運用を目指して、国立情報学研究所および企業と連携して進めていく。

WebELSに関しては、これまでの実績を基盤に、本研究が、大学院、企業、団体等における教育・研修ツールとして国内的・国際的にグローバルスタンダードの一翼を担う汎用e-Learningシステムとして普及するための事業化を目指す。このために、特許申請をベースとして(株)COMETを通してJSTの革新的ベンチャー活用開発に応募する（応募は20年度）など、安定的研究開発・事業化基盤を構築する努力を行うとともに、日本学術会議東アジア化学イニシアティブ分科会グローバル複素大学コンソーシアム（GUC）検討グループとの連携を深め、東アジアのMOU提携校との連携を推進する。さらに、UNESCOアジア本部との協力関係を強化すると共に、オーム社や協力企業、NPO 法人日本教育振興協会との連携によって、コンテンツビジネスや技術サポートビジネスを立ち上げ、コンソーシアム設立を図る等の整備を行う。

平成22年度以降の展開

この新領域融合プロジェクトでは、生命システム、地球システムを具体的な対象として、大規

模で複雑な巨大システムの解明に向けて、「生命」、「地球」、「統計数理」、「情報」の4分野横断型の融合研究を加速的に推進するための情報空間・情報基盤を確立することに取り組んできた。この間、情報の爆発や洪水は、その勢いをますます加速する傾向にある。すなわち、センサ技術や計測技術、観測技術の飛躍的な進歩により、「いつでも」「どこでも」「なにからでも」デジタルデータを取ることができる。さらに、高速大容量のネットワーク技術や記録・読み出し技術によって、時間や場所に依存することなく利用することができる。

また、携帯電話をはじめとするウェアラブル機器の発達によって、人間が、手軽に各種センサを身に着けることが可能となり、人間自身をセンサと見ることができるよう状況が生まれた。人間が「見る」「聞く」などにより感じる五感情報を伝えれば、多くの情報をキャッチする人間センサと見なすことができる。研究上の具体例をとれば、フィールド・リサーチなどが、これに相当する。

このように、現在、研究対象としている情報空間に比べて、データの多様性が桁違いに増すとともに、空間のスケールが大きく異なるものを同時に取りあつかう必要が生じる。さらに、文字、音声、音響、画像、映像などの異種データを意識することなく利用できなければならない。すなわち、マルチスケール・マルチモーダルなデータを取りあつかう「データ中心科学研究(Data Centric Science)の新しい手法に挑戦する必要がある。このようなデータ中心科学の手法により、「量と質」の異なる大量のデータから本質を導き出すことを目的とする。

平成22年度以降の出発点となるのは、本研究プロジェクトの成果である異種情報の結合・分類手法、例えば、「想・IMAGINE」を、データ中心科学研究の立場から、より実用性の高い汎用技術とすること、「論文」や「研究プロジェクト」を対象を拡大して、テキスト理解を試みる課題などがある。また、個別の環境適応に対応したゲノムの構造変化や、共生関係、極限環境生物集団としての全体的なメタゲノム解析など、これまでにない新しい知見や遺伝子資源が得られることが期待されたため、これらの資源を使った課題や、地球の全体像の把握を助ける、「Whole Earth」とでも呼ぶべきシステムを対象としたデータ中心科学的な研究が必要となる。

4. 研究費の推移

平成17年度実績： 176,130 千円

平成18年度実績： 156,170 千円

平成19年度見込： 152,370 千円

5. 平成19年度の研究推進体制

(1) 大規模・異種情報の収集・解析・結合・分類の手法および知識基盤の構築

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 佐藤真一 丸川雄三 相澤彰子 根岸正光
安達淳 大山敬三 孫媛 西澤正己 高須淳宏 市瀬龍太郎 柿沼澄男
[国立遺伝学研究所] 大久保公策
[統計数理研究所] 馬場康維 石黒真木夫 土屋隆裕 清水信夫
水田正弘（北海道大学，統計数理研究所客員教授）
[情報システム研究機構] 高久雅生

(2) 地球・生命などの巨大システム解明のための統合的情報基盤の形成と活用手法の確立

研究代表者

[国立情報学研究所] 藤山秋佐夫

共同研究者

[国立情報学研究所] 北本朝展 武田英明 市瀬龍太郎 佐藤真一 孟洋
[国立遺伝学研究所] 菅原秀明 城石俊彦
[統計数理研究所] 樋口知之 上野玄太 尾形良彦 村田泰章 種村正美
石黒真木夫 島谷健一郎
[国立極地研究所] 神田啓史 野木義史 土井浩一郎 宮岡宏 本山秀明
[新領域融合研究センター] 小林悟志 ムリアディ・ヘンドリー（4－9月）
川本祥子（4－12月）
[東京理科大学薬学部] 宮崎智
[名古屋大学] 井手一郎
[東京大学] 佐藤薫
[Johns Hopkins 大学] 大谷晋

(3) コラボレーションとコミュニティ形成のための情報共有基盤とバーチャルラボの構築

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] 藤山秋佐夫 上野晴樹 佐藤博之 何政
マフズル・ラーマン
[国立遺伝学研究所] 菅原秀明 阿部貴志 嶋本伸雄 富川宗博 佐々木裕之
桂勲
[統計数理研究所] 田村義保
[国立極地研究所] 岡田雅樹
[情報・システム研究機構] 榊川竜治
[総合研究大学院大学] 高畑尚之 [チュラロンコン大学] タワ・クワンパチュア

〔(株) オーム社〕 森正樹

〔清華大学 (中国)〕 張涛

〔メタメディア・テクノロジー (タイ)〕 ウッチチャイ・アンポー・ナランベス

〔(株) ゼネテック〕 岡野英司

6. 平成19年度の研究進捗

サブテーマ1では、異種情報の結合・分類手法の研究と大規模リンケージ情報の研究を具体的な課題として取り上げた。

異種情報の結合・分類手法の研究においては、新聞記事、環境問題資料、失敗知識情報、論文情報、専門辞典などの異なる情報源をユーザの興味によって関連づけて提示する利用環境のプロトタイプを開発した。各情報源を連想計算のウェブサービス付きで準備して、それらを動的に組み合わせるカスタムな情報源を構築する方式について検討し、「想・IMAGINE」システムを開発した。このシステムは、専門性や特徴の異なる情報源を横に並べて、それら全体を一つの情報源のように扱って、自由な文章を与えて類似な内容を収集する連想検索や、選択された検索結果の一部から連想検索を起動できる。知識のないユーザでも、最初は入門的な情報源から関連情報を収集し、そこからさらに深く掘り下げたいテーマを選択して再連想検索することにより、理解しながら専門的で正確な情報を得られる。また、研究者間のコミュニケーションがますます困難になりつつある生命科学分野で、困難さの原因が固有名称の多用や遺伝子の機能構造に関する自然言語表現にあることに注目して、それらを自動的にアイコン化して分野間のギャップを埋めるジーンアイコン（遺伝子象形文字）について検討した。

情報要素間の「つながり（リンケージ）」の収集・解析・活用手法の研究においては、学術的な情報に焦点をあて、特にリンケージを扱う上でポイントとなる研究者や研究機関の名寄せについて、参照用正解データの作成、共通要素数の統計的推定、自動同定手法の開発や同定サーバ実装等に関する研究を進めた。また、リンケージ情報分析の取り組みとして、計量書誌学的データに基づく科学分野の構造分析を行った。具体的には、「研究者情報サーバ」プロトタイプ版の構築と実運用への技術協力を行った。すなわち、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サーバの構築を進めるとともに、データベースやウェブ情報統合の基盤となるリンケージ要素技術の研究を行った。さらに、統計分野研究者データセットの作成し、大規模集団の名寄せの際の計画立案に資するため、複数団体からの無作為標本によるマッチングから、複数団体の共通メンバー数を推定する方法について理論的な検討を行った。また、論文データに基づく産学連携ネットワーク分析に関する方法論の検討を行った。さらに、情報および統計両分野の関連研究者の交流のため研究会を主催した。

サブテーマ2では、ライフサイエンスメタデータベースシステム構築の研究、極限環境生物統合データベースの構築の研究、および、地球環境データ統合データベースの研究を具体的な課題に取り上げた。

ライフサイエンスメタデータベースシステム構築の研究では、利用者主導型メタデータ DB システム構築を「新世代バイオポータル」プロジェクトの成果展開として推進する計画とし、データ量の増大と多様化に継続的な対処が可能なシステムの基本設計と、それに必要なハードウェアの整備を実施した。また、システム構築のためのデータ登録作業を続行するとともに、登録済みデータに対して機械およびエキスパートによる評価を実施した。また、PubMed Central 全文論文データベースを対象に、データベースを利用した論文、利用されたデータベースの種類、利用方法を抽出し、引用頻度上位100位までを高頻度利用データベースと定義して、エキスパートによるアノテーションと共に、メタDBに登録した。さらに、登録したデータのアノテーション作業を続行した。新たに改良した新パトロールソフトウェアにより、継時的に各DBのアップデート状況を監視するようにした。ここまでの作業で、概ね当初に計画した機能が実現できたため、インタフェースのリデザイン、バイオポータル辞書を用いた多言語検索機能を導入した強化タイプの実装を、日本語バイオポータルの改訂と合わせて年度内に行う予定である。

極限環境生物統合データベースの構築の研究では、国立極地研究所が収蔵しているコケ類資料を中心に、高精度3次元画像、ゲノム情報、分子進化情報等を基礎情報に加えて統合データベース化し、生物多様性についての新しい研究基盤を構築する計画とし、顕微鏡撮影した画像を3次元表示させるためのソフトウェア仕様の設定と、撮影用機材の整備開発を行った。極限環境生物統合データベースに用いる3次元画像データベース構築に必要なソフトウェアの改良とデータ収集を進め、web公開用のプロトタイプ(Koke3D)を作成した。Koke3Dソフトウェアについては、version10から11に改訂した。現在の仕様では、MacOS、WindowsXP 計算機のJava環境で稼働しており、対象物体は経度方向0-360°、緯度方向0-90°で自由回転とズームが可能である。各画像には、標本ごとの資料情報も同時に登録されていることから、例えばデジタル博物館DB等への応用も考えられる。画質を落としたWEB版も稼働状態にある。一方、南極由来コケ類試料のゲノムDNA解析については、試料側および実験手法側の双方に問題点のあることが分かり、次年度以降に向けてアプローチを根本的に再考しているところである。

地球環境データ統合データベース研究では、本プロジェクトにおける基本的コンセプトを「鉛直統合」と「水平統合」に明確化し、構築するポータルサイトを「鉛直地球(Vertical Earth)」と命名した。地球科学研究者との研究においても、地球科学データを「○○圏」という単位で統合するコンセプトに賛同を得た後、地球環境ポータル「鉛直地球(Vertical Earth)」のプロトタイプを公開可能な状態にまで整備し、ウェブサイト(<http://earth.nii.ac.jp/>)を2007年6月にオープンしてデータの公開を始めた。19年度までに公開したのは、気象庁のGPV(Grid Point Value)データを対象とするGPV Navigatorであり、全球のデータ(GSM)および日本付近のデータ(MSM)について過去数年分のデータを鉛直(上下)方向と時間方向に自由に行き来して閲覧可能なインタフェースを構築した。その他各圏のデータベースとしては「デジタル台風」および「台風前線」の機能増強を継続した。

南極域の地理情報システム研究では、まずシステムに登録すべきデータをリストアップし、優先順位の高い地形図や航空写真、衛星画像などからシステムに登録するとともに、「南極GIS」の

プロトタイプに関する極地研内の講習会とアンケートを実施し、南極に関する基礎的な地図データ等を蓄積した試行版を開発し、極地研究所内での利用に限定して公開した。南極 GIS サーバの保守・調整を継続して行い、GIS 上で基礎となる地図データの整備を行い、データの更新を実施した上で、地質図やその他のデータを GIS へ組み込んだ。特に、地磁気異常図に関しては、国際協力により南緯 60 度以南の地磁気異常図を公開した。

本サブテーマ 3 では、共同研究基盤としての「バーチャルラボ」構築基本ソフト NetCommons、および、高等教育向けマルチメディア教材共有型 e-Learning システム WebELS (Web-based e-Learning System) の研究開発を具体的な課題として取り上げた。

NetCommons は、研究者の分散環境において、距離の隔たりを感じさせない研究協調が可能なシステムの研究開発を目指し、NetCommons1.0, 1.1, 2.0 (α) の各バージョンをオープンソース公開した。また、国立極地研究所において、NetCommons を一斉導入し、運用上の問題点・改善点を検討した。特に、各研究室の情報公開用 Web 基盤としての活用が進み、第 49 次南極観測隊と関連研究機関の間の共同研究基盤として提供された。各地の教育委員会との連携を強化し、教育機関向けの情報共有基盤システムとしての定着を図った。これらの取り組みが評価され、総合研究大学院大学の葉山高等研究センタープロジェクトでは大学院教育の基盤として採用されたのをはじめとして、全国で 1500 以上の教育研究機関に NetCommons が導入された。また、多言語化の一環として、タイ語化および中国語化などのアジア言語化を進めた。

WebELS は、平成 18 年度より総合研究大学院大学の全学的なテーラーメイド教育推進、e-Learning 教育のプラットフォームとして導入され全学的に利用されることになり、コンテンツ開発と共に実利用の視点からの評価や改良を行ってきた。具体的には、非専門家用インタフェースの改良によるマニュアルレス化、同時アクセス処理機能の改良、多点 Internet 会議システムの安定化、同独自音声ラインの付加、オフラインビューアの開発等を行うと共に、GNU GPL 準拠のサービスを開始した。また、清華大、チュラロンコン大、ダッカ大や UNESCO 等との国際連携を推進しつつある。更に、日本学術会議東アジア化学イニシャティブ分科会に設置されたグローバル複素大学コンソーシアム (GUC) 検討グループに基盤ソフトとして協力することとなった。

7. 平成 19 年度の研究成果

サブテーマ 1～3 の各サブテーマの研究成果は、それぞれのサブテーマの研究報告書に明記する。