

GeNii－NII の学術コンテンツ・ポータルについて

国立情報学研究所開発・事業部 尾城 孝一

1. はじめに

国立情報学研究所（NII）は、前身の学術情報センターの時代から約 20 年にわたり、NACSIS-Webcat（総合目録データベース検索サービス）、NACSIS-IR（情報検索サービス）、NACSIS-ELS（電子図書館サービス）といったさまざまな学術情報提供サービスを提供してきた。しかしながら、それぞれのサービスは独立して存在しており、サービス間の連携は不十分であった。また、各サービスが固有の検索インターフェイスを備えていたこともあって、利用者にとっては必ずしも使い勝手の良いシステムとは言えなかった。

こうした限界を克服するために、既存の学術コンテンツを統合し、ひとつの窓口から多種多様な情報にアクセスできる環境を整備し、これを GeNii（NII 学術コンテンツ・ポータル）と名づけ、平成 17 年 4 月から正式にサービスを開始した。

2. GeNii サービスのコンセプト

GeNii の基本コンセプトを一言で表すと、“One for All, All for One”という言葉で表現することができよう。最近、インターネット上にない情報は存在しないも同然であるとみなす利用者が若い年代を中心に増えつつあると言われている。しかしながら、その一方で、インターネットのどこにどのような学術情報が存在しているかを的確に把握するのは容易ではない。この問題を解決するためには、以下のような作業が必要とされる。

- ・ 論文本体などの一次情報を可能な限り電子化すること
- ・ 一次情報を探すための二次情報、あるいはメタデータを整備すること
- ・ これまでに電子化し、蓄積してきた情報を、使いやすいように再構築すること
- ・ 今後増え続ける情報を効率的に収集し、組織

化すること

このような作業を通じて、個々の情報（One）を全ての利用者（All）に提供し、同時に、全ての情報（All）をその情報を必要とする利用者（One）に確実に届けること、それが GeNii 開発の際の中核的なコンセプトであった。

現在、GeNii は以下の 4 種類のサービスから構成されている。

- (1) 学術雑誌等の論文情報を集めた CiNii
- (2) 図書や雑誌の目録情報を集めた Webcat Plus
- (3) 科学研究費補助金による研究成果情報を集めた KAKEN
- (4) 分野別の専門的データベースの集合体である NII-DBR

さらに、これら 4 つのサービスを横断的に検索するための「総合検索」を加えたサービスが GeNii の全体像ということになる。

GeNii のトップページに現れる総合検索は、GeNii のサービス全体の中に、「どこに」、「どのような情報が」、「どのくらい」存在しているかを俯瞰するための検索インターフェイスである。一方、トップページからナビゲートされるそれぞれのサービス・コンポーネントは、コンテンツの特徴に応じて、データを検索し、表示するための最適なインターフェイスを備えている。

4 つのサービスの中で、CiNii だけは一部有料のサービスを含んでいるが、他のサービスは誰もが無料で利用できる。

3. サービス・コンポーネント

3. 1 CiNii (NII 論文情報ナビゲータ)

CiNii は、国内の学術論文を中心とした論文情報を提供するサービスであり、現在合わせて約 1,000 万件の論文情報が収録されている。CiNii の特徴としては、次の 3 点を挙げることができる。

(1) 引用・被引用リンク

CiNii には、ある論文が引用している論文、あるいは逆にある論文を引用した論文の情報が含まれている。利用者はこの引用・被引用のリンクを

たどって関連する論文情報を効率的に収集することが可能となる。

(2) 論文本体の提供

CiNii には論文の書誌情報（メタデータ）だけでなく、論文本体（PDF ファイル）の情報も収録されている。本体が利用できる論文数は、270 万件に達している。

(3) 網羅的な論文コレクション

CiNii には各種論文情報データベースの情報が集約されている。その核となるデータは、以下のとおりである。

・ NII-ELS（電子図書館サービス）（250 万件）

NII が国内の学協会の協力を得て作成しているデータベース。学協会発行の雑誌に掲載された論文の本体を画像データ（PDF ファイル）化して参照可能にしている。

・ 引用文献索引データベース（書誌 110 万件，引用 1,180 万件）

主として自然科学分野の雑誌論文に関して、ある論文がどの論文を引用しているかという情報を集録している。いわは SCI（Science Citation Index）の日本版をめざしたデータベース。

・ 研究紀要ポータル（書誌 70 万件，本文あり 20 万件）

大学等が刊行する研究紀要について、各刊行機関が作成したデータを NII が集積する。本体 PDF の収録も行っている。

・ 雑誌記事索引（690 万件）

国立国会図書館が収集する国内刊行の雑誌のうち、学術誌，大学研究紀要，専門誌を中心として、あらゆる分野の記事に関するデータを収録した国内最大の記事索引データベース。

以上の複数のデータベースには、同一の論文の情報が重複して収録されていることがある。CiNii には、同一の書誌と判定される論文書誌情報はひとつのレコードに統合して表示するシステムが取り入れられている。

CiNii のサービスは、無料で利用できる部分と有料の部分から成り立っている。論文の検索から、

検索結果の簡略表示までは誰もが無料で利用できる。また、出版元の学協会が許可している場合には、論文本体まで無料で閲覧することができる。

一方、抄録や引用情報を含む詳細書誌情報の表示や、学協会が課金の設定を行っている論文本体の閲覧は有料サービスである。有料サービスには、個人登録利用と機関定額利用(サイトライセンス)の2種類が用意されている。

3. 2 Webcat Plus (図書情報ナビゲータ)

Webcat Plus に収録されているデータの基礎を形成しているのは、全国の大学図書館等が NACSIS-CAT を使って共同構築している総合目録データベースである。現在、図書の書誌レコードが 780 万件、所蔵レコードが 8,400 万件収録されている。さらに、雑誌の書誌レコードが 30 万件、所蔵レコードは 420 万件に達しており、世界でも有数の目録データベースとなっている。このデータベースを一般公開するサービスとして、Webcat が既に存在していたが、その内容及び機能をさらに拡張し、日本の本探しのポータルサイトとして一層発展させたサービスが Webcat Plus である。

Webcat Plus には、以下のような内容や機能が「プラス」されている。

(1) 図書の目次・内容情報

書誌・所蔵データに加え、図書の目次・内容情報(紹介文)が追加されている。(株)トーハン、日本出版販売(株)、日外アソシエーツ(株)、(株)紀伊國屋書店の4社が著作権を有する図書情報データベースである「BOOK」には、1986年以降に発行された図書について、目次や、帯・カバー等に書かれている内容情報が収録されている。また、Nielsen Book Data には、現在入手可能、あるいは24ヶ月以内に出版予定である英語圏の英語図書について、目次等に書かれている内容情報が収録されている。さらに、幾つかの出版者から提供された出版目録データや出版者ホームページ等に掲載される新刊データ等の内容情報を含んでいる。

(2) 大学図書館等の未所蔵の資料

外部データを導入し、NACSIS-CAT の総合目録

を拡張している。これにより，大学図書館等の所蔵資料だけでなく，まだ所蔵されていない市販図書なども合わせて検索することができる。

(3) 連想検索機能

通常のキーワード検索の他に，連想検索というユニークな検索の仕組みを取り入れている。目的に応じて，従来型のキーワード検索と使い分けることにより，本探しを強力にサポートする。連想検索というのは，入力されたキーワードから関連性の高い単語をシステムが自動的に生成して，その関連語を含む図書をもれなく探し出す検索技術である。検索を行う際，的確なキーワードを選択することはそれ程容易なことではないが，連想検索を使用することによって，曖昧な検索語しか思いつかなくとも，自分が探している情報を効率的に見つけ出すことができる。

3. 3 KAKEN (科学研究費成果公開サービス)

文部科学省と日本学術振興会が交付している科学研究費補助金によって行われた研究に関する情報を集約したサービスが KAKEN である。NII では，これまでも NACSIS-IR を通じて，各年度に科研費の交付が決定された研究課題のデータベースと，交付を受けて行われた研究の実績や成果の概要を収録するデータベースを提供してきた。しかしながら，交付決定情報と交付後の成果情報が別々のデータベースになっており，使い勝手の点で問題があった。それを整理，統合し，採択課題毎に，実績報告，成果概要を一覧することができるようにエンハンスした。また，複数年度にまたがって研究が行われている場合は，年を追って実績報告や成果概要を閲覧することができる。

3. 4 NII-DBR (学術研究データベース・リポジトリ)

NII-DBR には，国内の学会，研究者，あるいは図書館が作成した専門的なデータベースが収録されている。こうしたデータベースは，それぞれの研究分野に特化しており，専門性の高い貴重な学術コンテンツのひとつである。しかしながら，こうしたデータベースを維持管理していくには，相当のコストを必要とする。また，年を経るにつれ

て、メンテナンスの担当者が不在となるケースも起こりうる。NIIでは、こうした貴重なデータベースを積極的に受け入れて、公開していきたいと考えている。現在、27のデータベースを受け入れており、合わせて150万件のレコードが検索できる。

4. 今後の展望

4.1 利用者志向のサービスに向けて

最近の電子情報環境の変化は急激であり、利用者である研究者や学生のニーズも多様化し、かつ日々変化している。学術コンテンツ・サービスも、こうしたエンドユーザの需要に的確に応える努力を怠るやいなや、瞬く間に陳腐化するという運命に遭遇する。

NIIでは、GeNiiの正式リリース後1年余りが経過した現在、利用者の声に広く耳を傾けるためのアンケート調査を開始した。具体的には、GeNiiの各コンポーネントのトップページに、質問回答のフォームを用意し、そこから利用者の声を広く聴取したいと考えている。まず手始めに、Webcat Plusに関するアンケート調査をこの7月から開始した。準備が出来次第、CiNii, KAKEN, NII-DBRへと範囲を拡大していきたいと考える。

こうして集約したエンドユーザの意見を整理、分析し、次の世代のGeNiiのサービス設計に活用していくことが今後の大きな課題のひとつである。

4.2 外部情報サービスとの連携

GeNiiは学術情報への総合的窓口をめざしたポータルサービスである。利用者はこのポータルを通じて多種多様な学術コンテンツを効率的に発見し、アクセスすることができる。しかしながら、現在のインターネット上には、GeNiiの他にもさまざまな学術的なポータルサイトが存在している。例えば、医学関係の文献データベースである医中誌Webや科学技術文献の統合的な検索サイトであるJDreamIIなどが国内の代表的な学術系ポータルサイトである。また、Google, Yahoo, さらにマイクロソフトといった民間の情報サービスも学術情報への興味を示している。

これらのポータルは、自らの存在感を高めるべく、サービスの相互乗り入れに向けた取り組みを強めている。こうした動きの中で、外部サービスとのリンクの扉を閉ざすことは、インターネットの大海の孤島となることを意味している。

GeNii も遅ればせながら、外部サービスとの連携に着手した。その嚆矢として、現在 CiNii と Google Scholar とのリンク・システムを構築中である。このシステムの平成 18 年秋のリリースをめざしているが、これによって、Google Scholar を通じて、CiNii に収録されている国内の学術論文の検索が可能となる。さらに、利用者は、Google Scholar のリンクを通じて CiNii 画面にナビゲートされ、そこから論文本体にアクセスすることが可能となる。この連携により、Google Scholar は国内の学術論文情報入手することが可能となる。一方、CiNii は Google Scholar を通じた集客力の向上によって、論文本体の可視性を高めることができる。いわば両者の間には win-win の関係が確立されることになる。

また、医中誌 Web と CiNii の相互リンクの計画も進行中である。医学中央雑誌からメタデータの提供を受け、現在それを CiNii のメタデータと同定し、統合する作業を進めている。同定の結果は、医中誌 Web にも反映される。この連携が完成すれば、医中誌 Web と CiNii との間で、相互リンクが実現することになる。このサービスも平成 18 年の秋にはリリースしたいと考えている。

さらに、外部サービスとの連携という観点からは、さまざまな情報ベンダーが提供しているリンクリゾルバや横断検索サービス（Federated Search）への対応も欠くことができない。今後の重要な課題のひとつである。

4. 3 研究成果の反映

そもそも NII の使命は、情報学に関する総合的な研究・開発を行う「研究教育」と、学術情報の流通のための基盤整備を行う「事業」の 2 つから成り立っており、この 2 つをいわば車の両輪として有機的に連携していく点が NII の大きな特長となっている。こうした NII ならではのメリットを

活かすべく、現行のコンテンツ事業のチェック・アンド・レビューと次世代の学術コンテンツ・サービスの研究・開発を目的として、所内の研究系職員と事務系職員とから構成される学術コンテンツ・サービス研究開発センターが平成 18 年 4 月に設立された。

今後は、センターの活動から生まれた新機能を順次 GeNii のサービスに組み込んでいきたいと考えている。センターの当面の研究課題として、Web2.0 的な要素を取り入れた次世代 CiNii の開発を挙げることができる。今年度にプロトタイプ版の開発を行い、可能な限り早期にその成果を現在の CiNii に付加していきたいと考えている。また、CiNii で提供している論文本体の PDF ファイルを OCR 処理により全文テキスト化し、それを元にして、付加的な価値を伴う全文検索サービスを展開するというテーマも今年度の課題に挙げられている。

5. おわりに

NII では、現在、最先端学術情報基盤 (Cyber Science Infrastructure: CSI) 構築の一環として、学術コミュニティにとって不可欠なコンテンツを、大学等との密接な連携により形成、確保し、付加価値を付けて広く発信するための次世代学術コンテンツ基盤整備事業を多角的に進めている。この事業により形成された多様な学術コンテンツへの統合的利用窓口としての機能を果たすことが GeNii の主たる役割である。NII は、今後も利用者の声や時代の変化に鋭敏に反応し、真の意味での学術コンテンツ・ポータルへと GeNii を進化させていきたいと考えている。

参考 URL :

GeNii. <http://ge.nii.ac.jp/>

CiNii. <http://ci.nii.ac.jp/>

Webcat Plus. <http://webcatplus.nii.ac.jp/>

KAKEN. <http://seika.nii.ac.jp/>

NII-DBR. <http://dbr.nii.ac.jp/>

(本文中のデータ件数は、全て平成 18 年 6 月時点の数字である。)

GeNii: NII Academic and Scholarly Information Portal
(By Koichi Ojira, National Institute of Informatics)

National Institute of Informatics (NII) has provided a wide range of scholarly information services such as NACSIS-Webcat, NACSIS-IR and NACSIS-ELS nearly for twenty years. But each service was isolated and linking system between services is not very effective. So we integrated those services into the one-stop portal called GeNii. GeNii was launched officially in April 2005. It has grown to the flag-ship content service of NII in less than a year. In this article the services provided by GeNii are outlined. Also the future prospects including cooperation with other information vendors are discussed.