

CA1691 XXXXXXXXXX 学術情報プラットフォームとしての CiNii

1. はじめに

国立情報学研究所（以下、NII）が運営する論文情報ナビゲータ CiNii（サイニイ）は、サービス開始から5年目を迎えた2009年4月に新しいシステムの導入を行った（図1）。

筆者が携わった今回のシステム導入では、学術情報プラットフォームとしての CiNii の立脚点を明確にし、それに沿った CiNii の再設計・開発を行った。本稿では、システム導入の概要について述べるとともに、その背景にある学術情報流通プラットフォームの考え方について概説する。



図1 CiNiiのスクリーンショット

2. CiNii の概要

CiNii は学術論文を対象とした国内最大級の情報サービスである。NII が電子化している約 300 万件の学協会誌・大学研究紀要だけでなく、国立国会図書館の雑誌記事索引、科学技術振興機構の J-STAGE、Journal@rchive、ならびに各大学・研究機関で構築が進む機関リポジトリと連携し、計 1,200 万以上の書誌データを検索することができる。また、NII が構築している引用文献索引データベースに基づき、論文の引用・被引用関係を表示することができる。

CiNii は 2005 年 4 月から提供されており、現在は 5 年目に入ったところである。この間に、学術情報サービスは「ウェブ 2.0」という言葉に代表される情報環境の激変に直面することになった。ユーザー層が格段に広がり、専門家・学生に対するサービスだけでなく、一般ユーザーにも目を向けた情報流通の基盤としての機能が求められるようになった。

こういった流れを受けて、CiNii では 2006 年 12 月

から 2007 年 4 月にかけて大幅なリニューアルを行った。このリニューアルでは、契約機関に所属している研究者・学生のみアクセス権限が与えられていた書誌情報を原則一般公開し、Google あるいは Google Scholar といった外部の検索エンジンから検索可能とした。リニューアルの結果、利用率を示す各種指標（検索回数・書誌情報表示回数等）において、3～10 倍程度の大幅な増加が記録されるとともに、多数の一般ユーザーの流入が確認された。このリニューアルの目的や内容に関する詳細、経過は拙論文に詳しい⁽¹⁾。参考にされたい。

3. 学術情報プラットフォームのために

3.1 課題

前述のリニューアルは大きな成果を上げたものの、一方で新たな課題を浮かび上がらせることにもなった。

ここでは主な課題として 3 点を挙げる。

- パフォーマンス

アクセス数の急激な増加は CiNii のシステム全体に極めて重大な影響を与えた。2005 年 4 月のサービス開始時、さらに遡れば設計時の想定をはるかに上回る数のアクセスを処理しなければならず、結果としてスローダウンやシステム停止といった事象が頻発した。こういった事象には可及的速やかに対応しなければならないが、システムの特長によっては改修に多大な時間・コストを必要とする場合があることから、どのような方法が適切かを見極めるのは難しい。

- ユーザビリティ

外部の検索エンジンとの連携によって、新たなユーザーが CiNii を訪れる機会が増加した。そのうちの多くは専門家ではない一般ユーザーであると思われる。これら一般ユーザーにとって、専門家向けに設計された検索インターフェイスや書誌詳細表示画面はなじみのあるものではない。その結果、論文情報を有効活用しないまま CiNii から立ち去るといった例が多く見られた。専門家と一般ユーザーが学術情報サービスに望むものは大きく異なるものと予想されるが、両者が一定の満足度を得られるユーザービリティを実現することは急務である。

- オープン化

前 2 点とは異なり、目の前にある課題ではないものの、多くのウェブサービスで取り入れられつつあ

るオープン化の動きをどのように取り入れるかは CiNii の将来を考える上で重要な問題である。ここでのオープン化とは、情報を単純に一般公開することではなく、機械処理されることを念頭に置いた構造化データの配信・配布を意味する。

これら3点の課題は、CiNii が学術情報プラットフォームとして機能するかどうかの試金石であるといえる。

自らがプラットフォームと名乗れるようになるためには、多数のアクセスにも耐えられる設計であること、あらゆるユーザに開かれた使いやすいシステムであること、また他のサービスがプログラムを通じて一部の機能を組み込めるようにすること、の3点は必要条件であると言って差し支えない。また機械処理を許可するためには、大量のアクセスを高速に処理し、意図しない遅延やダウンが発生しないような設計が求められる。

3.2 設計

新システムでは、上記の3点の課題を解決し、CiNii が学術情報プラットフォームとして機能することを最優先事項として設計・開発を行った。

まず、パフォーマンスについては、旧システムの機能分析を行い、データを作成するためのバックエンドシステムと作成されたデータに基づくサービスを行うためのフロントエンドシステムを完全に切り離し、フロントエンドシステムの性能向上に特化することとした。その上で、ごく少数の大型計算機で処理を行うスケールアップ・アーキテクチャを廃し、小型のサーバを必要に応じて追加することでパフォーマンスを向上させることができるスケールアウト・アーキテクチャを採用した。これによって、数年後の利用予測に左右されず必要最小限のコンパクトなシステムを構築できる状況を整えた。

次に、ユーザビリティについては、設計の最初段階からユーザビリティの専門家と協力し、ユーザが体験する画面遷移のモデル作成と個別のページの詳細なデザインを行った。この過程では筆者をはじめとする NII 担当者をユーザと見立てた徹底的なユーザ中心設計が行われた。また、これらの作業は一般的なシステム構築で必要となる機能面での要件定義とは完全に切り離れた状態で行い、ユーザビリティ設計の成果物としての画面遷移図ならびにデザインをそのままシステム構築のための仕様とすることで、ユーザビリティが重要な要素であることを関係

者に知らしめる効果があった。

オープン化への対応では、CiNii の主要な機能である検索機能と書誌詳細表示機能について、機械処理が可能なように構造化されたデータを入手できるように設計した。検索機能についてはデファクトスタンダードである OpenSearch に準拠し、検索結果一覧のリストを RSS 1.0 形式ないし Atom 1.0 形式で入手することができる。また、書誌詳細表示機能では、RDF (Resource Description Framework) に則った書誌データの表現を行っている。いずれも書誌データならではの属性を表現するために Dublin Core や PRISM (Publishing Requirements for Industry Standard Metadata)⁽²⁾ といった標準規格団体が提供する語彙を使用している。これらは、データを活用する側である開発者の負担を可能な限り小さくしたいと考えているからである。また、OpenSearch の返戻形式の一方である RSS 1.0 と書誌で用いられる RDF はデータ記述方法が統一されており、また相互にリンク付けされているため、機械処理によって検索から論文の発見、著者名や所属名の抽出までが容易に行える。これは、機械処理によってウェブ上のデータを知的に活用するセマンティックウェブの理想に近い。

検索の OpenSearch ならびに書誌 RDF は、コンピュータからのリクエストを受け付け、それに適したデータを返戻するという意味でウェブ API の一種であるといえる。筆者らは、CiNii においてパフォーマンス強化の施策を行ったことをふまえて、このウェブ API を広く普及させ、開発者コミュニティの育成を行いたいと考えている。その先鞭をつける試みとして、2009年6月から9月にかけて CiNii ウェブ API コンテストを開催し、これらのウェブ API を利用したアプリケーションを募集している⁽³⁾。

4. 成果と今後

2009年7月現在、新システムの導入から約4か月が経過しているが、その効果は非常に大きい。前回のリニューアル後の2007年4月から2009年6月までの月間の検索回数ならびに月間の本文 PDF ファイルダウンロード回数を図2に示す。導入直後の4月において検索回数は前年同期比1.8倍、本文 PDF ダウンロード回数は2倍程度の伸びが見られる。また、これらの数値は過去のすべての期間における最高値であり、5月・6月と月を追うごとにさらに増加している。

CiNiiの概況 (利用傾向)

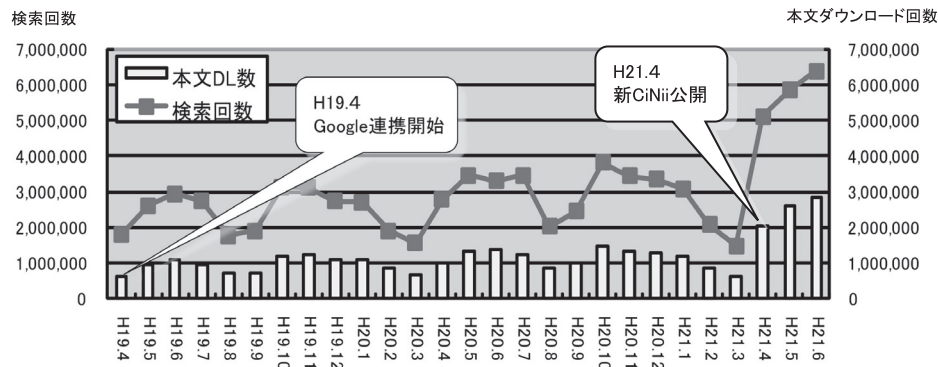


図2 リニューアル後の利用回数の推移

アクセス増加の要因にはさまざまなものが考えられるが、システムの高速化とユーザビリティの向上によって、ユーザがCiNiiの中で何度も試行錯誤しながら検索しているのではないかと推測される。また、本文PDFファイルダウンロードも順調に増加していることから、最終的には満足のいく検索結果が得られていることがわかる。今後はウェブAPI経由のアクセスが増えると予想される。機械的なアクセスを制御することは難しいが、開発者と協力しながら互いにとって実りのある関係を築けるよう、システムの監視やチューニングを継続的に行っていく所存である。

5. おわりに

筆者らは学術情報プラットフォームの実現を目指して、論文情報ナビゲータCiNiiの再設計、開発を行ってきた。現段階ではパフォーマンス等に問題はなく、頑健なプラットフォームを作ることができたのではないかと自負している。今後はウェブAPIを通じて外部の開発者との連携を強化し、新たなアイデアを常に取り入れられる体制を構築する予定である。また、研究と事業を1つの組織で行っているNIIの特徴を生かし、研究成果を積極的に学術情報プラットフォームに取り入れていくことを目指す。

(国立情報学研究所：おおむかい一輝)

- (1) 大向一輝. 学術情報サービスのユーザモデルとファインダビリティ. 情報の科学と技術, 2008, 58(12), p. 595-601.
- (2) "PRISM". IDEAlliance.
http://www.idealliance.org/industry_resources/intelligent_content_informed_workflow/prism, (accessed 2009-08-10).
- (3) "CiNii ウェブAPI コンテスト 実施要項". CiNii.
http://ci.nii.ac.jp/info/ja/web_api_contest_2009.html, (参照 2009-08-10).