

国立国会図書館における研究開発（R&D）

—システムを中心に—

国立国会図書館 電子情報部 電子情報サービス課
次世代システム開発研究室長
原田 久義

1 次世代システム開発研究室の設置とその目的

次世代システム開発研究室（以下、「次世代室」といいます。）は 2011 年 10 月に実施した機構改革における、電子情報部の発足に伴い設置されました。設置の目的は、先進情報技術を応用した新しい図書館サービスを実現するための次世代システムに係る調査研究・実証実験を行うことにあります。次世代室の体制ですが、職員 3 名、非常勤調査員 1 名、館内の他部局の研究協力員 1 名、大学や研究機関の委嘱研究員 5 名となっています。この委嘱研究員との連携協力が次世代室の活動を支えています。設置から 6 年が経過しましたが、これまでの主な成果と今後の取組についてご紹介します。

2 標準化に関する調査研究—Linked Data 及び識別子に関する取組

NDL では、ウェブ世界における NDL のデータの相互運用性を向上するための取組を行ってきました。このうち、Linked Data 及び識別子に関する取組についてご紹介します。

まず、NDL の統合検索ポータル「国立国会図書館サーチ」で提供する書誌データは、2012 年の公開時から、当館が維持する日本国内の情報資源のメタデータ記述のための標準、「国立国会図書館ダブリンコアメタデータ記述（DC-NDL）」で記述され、Linked Data として取得可能にしています。

2014 年には、Linked Data に関する現況調査と有識者ヒアリングを行い、NDL の Linked Data の課題を整理しました。ここで抽出された課題を解決し、NDL 外のシステムから NDL のデータを使いやすくするために、提供する Linked Data を増やす取組、データ利活用イベントの開催、NDL の URI 体系の管理に着手しました。こちらは、2015 年に NDL が新たに LD として公開した図書館及び関連組織のための国際標準識別子（ISIL）を用いて、外部の方が作成したアプリです。ISIL を付与された日本の図書館や情報センターの位置や地域ごとの館数が分かります。

また、2017 年には識別子に関する現況調査を行いました。その結果は、DOI や

典拠 ID 等、NDL が付与している各種識別子の今後の先端的な活用方法の検討に資する予定です。

3 次世代システムに係る調査研究

(1) これまでの成果

図書館を取り巻く環境が急速に変化する中、当館においても、次世代図書館サービスの検討及び構築を進める必要があります。そのためには、次世代図書館サービスにとって有効な要素技術の検証が必須になります。

そのため、大学や民間の研究者と連携し、研究者の研究開発成果について、当館が構築しようとする次世代図書館サービスへの適用可能性を検証する場として、NDL ラボを設置し、2013 年 5 月にサイトを公開しました。

電子読書支援システムは、一定の文字情報を基に連想検索を行い、Wikipedia 等の他の情報源から関連する参考情報を候補として利用者に提示し、読書や調査の支援に資することを目的とするシステムです。検索対象となるコンテンツとして、当館のデジタル化資料から著作権保護期間が満了した資料を選定しました。国立情報学研究所の阿辺川武特任准教授の協力を得て実施しました。

NDL ラボサーチは、当館作成書誌（図書、雑誌、新聞、雑誌記事索引等）を検索対象とし、検索キーワードのオートサジェスト、検索結果に関連したタグクラウドによる絞り込み等の実験的な機能を追加したシステムです。検索機能の高度化を企画し試行するための環境として使用するプロトタイプで、2013 年 9 月に一般公開しました。

翻デジは、当館のデジタル化資料のうち、著作権保護期間満了の資料を対象として、クラウドソーシングにより本文テキストの共同入力・校正作業を行うプラットフォームです。当館のデジタル化資料は画像で提供しておりますので、その更なる活用策として、不特定多数の人々による本文テキストの入力・校正により、単語単位での検索に資する可能性を追求し、利用者の利便性を向上させることを目指しています。人文情報学研究所の永崎研宣主席研究員の協力を得て実施しました。

書誌情報検索・可視化システムは、美馬秀樹東京大学准教授を中心として開発された、文書間の意味的な関連に基づいてデータマイニングを可能とする検索・可視化システム「MIMA サーチ」を活用して構築しました。当館でデジタル化した図書の書誌データ約 92 万件及び目次データ約 71 万件を対象に検索・可視化を行います。

視覚障害者等へのテキスト化資料提供実験も行っています。社会福祉法人日本点字図書館と当館が、2015 年 3 月に覚書を締結し、視覚障害者等へテキストデータを提供するため、共同校正システムを用いた実験を実施しています。2015 年 4 月～6 月にかけて当館デジタル化資料から 5 タイトルを選定し、画像データからテキスト DAISY を製作した結果、テキスト DAISY、EPUB、プレーンテキストの製作に

同システムが利用可能であることを確認しました。

(2) 今後の課題・展望—2020年に向けて

今後の次世代室の活動ですが、今年の7月に計画を策定しました。その中に2020年までの4年間の具体的な取組事項を掲げていますのでご紹介します。

1点目は、多様な電子情報資源の所在情報の集約・統合及びコンテンツの利活用の促進として、分野横断型検索・ナビゲーションサービスの実現を目的とした、ジャパンサーチ（仮称）のプロトタイプ構築です。

2点目は、研究データの保存と利活用です。研究データに係る当館の関与の方向性を確認することを目的とした、研究データの保存・利活用に関する調査研究、また広報を目的としたイベントを実施します。

3点目は電子情報の長期利用保証です。パッケージ系電子出版物のマイグレーションを実施し、試行を踏まえて、実際の業務に反映できるよう事業化を目標としています。また、当館システムへの適用を視野に入れた、長期保存に係る要素技術の調査研究も行います。

4点目はAIの活用です。例えば書誌作成の効率化として、書誌データ入力補助を目的としたNDC分類の自動付与、さらにテキストデータを用いた書誌データの自動生成等の試みを考えています。テキストの構造化に係るディープラーニング等のAIに関する技術的な調査研究も行います。

また、AI等の技術を当館の業務プロセス・サービスへ適用して、利用者サービスの向上が図れるかどうかを検討・試行します。具体的には検索精度の向上及びレファレンス業務への適用可能性について検討します。

AIについて、中国は国としての取組を推進していると伺っています。このような機会にいろいろと学ばせていただければ幸いです。ご清聴ありがとうございました。