

NDL ラボの取り組みとデジタルヒューマニティーズ

川島隆徳^{†1} 丹治美玲^{†1}

国立国会図書館では、2013年からNDLラボという事業に取り組んでいる。この事業では、外部機関との協力により新しい図書館サービスを実証実験することを目的としており、いくつかの実験的システムが稼働している。このような「ラボ」事業は、海外でも事例が増えてきているが、その中にはデジタルヒューマニティーズと関わりが深いものも多い。本稿では、NDLラボ事業の内容と提供可能なデータ、デジタルヒューマニティーズ分野との更なる連携の可能性について述べる。

1. はじめに

情報技術の急速な進展の中、図書館においても従来の枠にとどまらない先進的なサービスの提供が必要とされている。国立国会図書館では、そのようなサービスを検討するためにNDLラボという取り組みを2013年から開始した（NDLはNational Diet Libraryの略）[1]。先進的なサービスの実現には調査研究が欠かせないが、国立国会図書館単独での取り組みにはスキルの限界がある。そこで、NDLラボでは大学や研究機関等の研究者に協力を求め、データやインフラを提供する代わりに実験的サービスシステムを提供して頂くという枠組みを取っている。

似たような取り組みは、現在世界各国で行われている。いずれの取り組みも、図書館が持っているデータの特性から、自然とデジタルヒューマニティーズと関わりのある取り組みとなっている。本稿ではそのような取り組みと、現在NDLラボで稼働している3つの実験的サービスについて述べる。

2. 各国での取り組み

「ラボ」と銘打つ、ないしは類似の目的意識を持った実験的事業は、海外の国立図書館・公共図書館等においても散見される。本節では、4つの事例を取り上げて紹介する。

2.1 British library Labs

British Library Labs[2]は大英図書館の実験事業である。Andrew W. Mellon財団の出資を受け、デジタルコレクションへのアクセス提供方法を変えることを目的とした2年間のプロジェクトとして2012年末から活動を開始している。

年に一度コンペティションを実施し、受賞者のアイデアを基に共同開発を行うほか、データセットやツールの紹介・開発やイベント開催等の活動を行っている。

データの例としては書誌データの他、スキャンされた資料の挿絵部分が写真共有サービスFlickrを利用して1,000,000点以上公開されている。コンペティションでは、例えばヴィクトリア朝時代のジョークを分析し、半自動で抽出した関連性のあるデジタル画像上に吹き出しで表示さ

せ、公開・共有できるツールなどが選ばれている。

2.2 HathiTrust Research Center

HathiTrust Research Center[3]は、研究者がHathiTrustの提供する資料の画像・全文テキストの一部を活用してテキストデータマイニングや分析を行うための環境の構築を目的とした事業である。イリノイ大学・インディアナ大学と提携し、2011年にプロジェクトが開始された。HTRCはプラットフォームを用意し、テキストマイニングツール提供等のサービスを行っている。利用には登録が必要である。

“Workset Builder”から資料を検索・選択してデータセットを作成し、必要に応じて整形した上で“Portal”にデータを登録し、分析手法を選択してジョブを実行することでテキストマイニングが行える。

他に、大量データ取得に適したHTRC APIの提供等を行っている。

2.3 Europeana Labs

Europeana Labs[4]はEuropeanaのコンテンツの二次利用促進を目的として設置された。Europeanaの2つのプロジェクト、Europeana Creative及びEuropeana Spaceによって運営される。Webサイト（2014年4月にβ版サイト公開）と物理的なラボ施設から構成され、Webサイトには、APIやデータセット例、アプリ等が紹介されている。2015年4月時点では、データセット例は79、紹介されているアプリの数は168件となっている。具体的なアプリケーションの例としては、画像の類似性を利用して徐々に検索を絞り込んでいく画像検索のシステムなどがある。

2.4 New York Public Library Labs

New York Public Library Labsは、2007年頃、デジタル環境への移行の中で、デジタル人文学分野の急激な進展に対してより良いポジションをとるための戦略計画プロセスの一環として、インターネット時代の図書館を再構成することを目的として設置された[5]。

図書館内でツール・アプリケーションを開発し、一般向けに実験的サービスを提供している。主な成果に、クラウドソーシングによって歴史地図と現代の地図のマッピングを行うアプリケーション“Map Warper”や、レストラン等の

^{†1} 国立国会図書館
National Diet Library

メニューコレクションの共同翻刻を行うアプリケーション”What’s on the menu?”の公開, ”New York Public Library Digital Collections”の API 作成・提供等がある。外部のアイデアを取り入れて作成したアプリケーションも存在する。独自に開発を行う以外に、他機関との連携協力も行っており、成果物がサイトで紹介されている。

3. NDL ラボでの取り組み

3.1 NDL ラボの枠組み

NDL ラボ[6]は、国立国会図書館による調査研究事業である。国立国会図書館が開発した実験的システム及び外部の研究者の実験的システムを公開している。現時点では、Europeana Labs 等とは異なり、無制限にデータを公開するのではなく、外部の研究者には国立国会図書館の委嘱研究員となつていただくことでデータとインフラを提供する形式となっている。なお、国立国会図書館では各種データの API による提供も積極的に行っているため、NDL ラボを利用することで特に利用可能なデータが増えるというわけではない。しかしながら、館内のインフラからは通常とは異なる条件で API の利用やデータ取得ができるため、実験システムの開発における利便性は向上する。

NDL ラボでは、他にもアイデアソン[7]を開催するなどの取り組みも行っている。

以降、NDL ラボの枠組みの上で稼働している3つの実験システムについて具体的に紹介する。

3.2 電子読書支援システム

電子読書支援システムは国立情報学研究所の阿辺川武特任准教授の開発した、デジタル化された書籍を便利に閲覧するためのシステムである。

対象となるデータは画像化された資料で、通常の電子書籍ビューアと同様にページめくり等の機能がある他、OCR 等によって資料中のテキストを認識し、そこに含まれているキーワードとその情報を Wikipedia 等の外部ソースから取得・表示する機能を持っている。

その他、抽出したキーワードを利用した索引の表示等も可能となっている。

3.3 翻デジ 2014

翻デジ 2014 は、人文情報学研究所の永崎研宣主席研究員及び日本デジタル・ヒューマニティーズ学会の分科会である SIG-Transcribe JP が提供するテキスト翻刻のための仕組みである。国立国会図書館がデジタル化した資料の中でもインターネット公開しているものは特に近代デジタルライブラリーというブランドで更改されているが、翻デジ 2014 はこの近代デジタルライブラリー上の資料を翻刻することを目的としている。

翻刻はユーザ登録によって誰でも行えるようになっており、いわゆるクラウドソーシングによるデジタル化の仕組みとなっている。

3.4 NDL ラボサーチ

NDL ラボサーチは、国立国会図書館サーチの将来の姿を検討するためのプロトタイプ検索システムである。とは言っても国立国会図書館サーチと比べてそれほど機能が増えているというわけでは無く、主に JavaScript を用いた実装で検索応答性能を向上させることを主眼に構築されている。また、実験的な機能として検索結果の書誌情報に含まれるキーワードからタグクラウドを表示したり、ブックマークした書誌から、分類番号などを利用して関連書誌を検索提示したりする機能を持っている。

現在は、検索ログからの類似書誌の抽出提示機能などを開発している（未公開）。

4. 利用可能なデータと環境

4.1 NDL ラボで提供しているデータ

NDL ラボで提供可能なデータは、著作権等の問題から館外に何らかの形で公開しているデータに限られる。具体的には、以下のデータが利用可能である。いずれのデータも API による取得の他、ファイルで一括して提供することも可能である（画像データを除く）

表 1 NDL ラボで利用可能なデータ

データ	件数
書誌データ	約 1106 万件
雑誌記事の索引データ	約 1133 万件
典拠データ	約 114 万件
著作権切れの資料の画像データ	約 48 万冊
資料の目次データ	約 131 万件

なお、書誌データ、雑誌記事の索引データについては国立国会図書館サーチ[8]から、典拠データについては Web NDLA[9]から、インターネットからでも API でデータの取得が可能である。画像データや目次データについては、国立国会図書館デジタルコレクション[10]から閲覧が可能である。

4.2 NDL ラボ環境

NDL ラボでは、実験システムを公開するための仮想サーバを提供している。また、各種のデータにアクセスする API や、画像に OCR をかけた結果を返す API なども用意している。委嘱研究員は、クラウドサービスのようにこれらのインフラを利用可能となる。

5. おわりに

5.1 デジタルヒューマニティーズとの関わり

NDL ラボ自体は特にデジタルヒューマニティーズを志向している訳では無く、NDL ラボサーチなどは単純に国立国会図書館の 1 システムの実験版に過ぎない。また、電子読書支援システムも、デジタル化資料をどのように閲覧してもらおうか、という観点のシステムであり、必ずしもデジ

タルヒューマニティーズと結びつきが深いわけではない。しかし、「翻デジ」は、動機の一つは近代デジタルライブラリーの資料をインターネットから検索可能にしたい、という一般的なものではあるものの、得られるテキストデータは近代のものとなり、仮に分析などを行うとしたら、それはデジタルヒューマニティーズの範疇に入ることになるだろう。

各国のラボ事業でも状況は類似しているが、デジタル化した資料のデータを利活用する、ということ考えるとそこには何らかの形で人文学が関わってくることになる。提供可能なデジタル化した資料は必然的に「古い」資料が多く、必ずしも一般的な図書館サービスの利用者像において要求のある資料とはならないからである。そうなってくると、図書館の外に研究資源を求める場合、協力していただけるのは情報処理の専門家や企業というよりも、そのような専門性を持ちながらも古い資料に興味を持っていただけるデジタルヒューマニティーズの研究者ということになる。

とはいえ、より一般的なニーズの資料を対象として同じシステムを利用することも可能であって、人文学のためだけの取り組みにはならないと考えている。デジタルヒューマニティーズにとっても、そのような仕組みが社会に還元されることになれば、学問の価値を社会に示すことができるのではないだろうか。

デジタルヒューマニティーズにおいても、特に分析のツールなどは人文学だけでなく広く一般に役立つことが想定される。例えば、一般的な書誌や全文検索の仕組みでは無く、専門家が利用する検索や、対象テキストの分析機能が組み込まれたシステムなどがあれば、一般的な利用者も専

門家の目でテキストを見ることができるようになる可能性なども考えられる。

5.2 NDL ラボの今後

NDL ラボでは、今後も実験システムを追加していく想定である。また、実験システムの成果を実際のサービスに還元していくことにも取り組んでいく。

利用可能なデータの拡充や、内部的な API の強化、複数システムの連携等にも取り組んでいきたいと考えているため、今後ともご支援をお願いしたい。

参考文献

- 1) 国立国会図書館: ようこそ、実験室へ NDL ラボの誕生・現在・未来, 国立国会図書館月報 No.640/641
http://dl.ndl.go.jp/view/download/digidepo_8704435_po_geppo14078.pdf?contentNo=1
- 2) British Library Labs
<http://labs.bl.uk/>
- 3) HathiTrust Research Center
<http://www.hathitrust.org/htrc>
- 4) Europeana Labs
<http://labs.europeana.eu/>
- 5) NYPL Labs
<http://www.nypl.org/collections/labs>
- 6) NDL ラボ
<http://lab.ndl.go.jp/>
- 7) 国立国会図書館のウェブページを使い尽くそうアイデアソン
<http://lab.ndl.go.jp/cms/?q=opendata2015>
- 8) 国立国会図書館サーチ API
<http://iss.ndl.go.jp/information/api/>
- 9) Web NDLA
<http://iss.ndl.go.jp/ndla/sparql/>
- 10) 国立国会図書館デジタルコレクション
<http://dl.ndl.go.jp/>