

## 国立国会図書館サーチと日中韓自動翻訳機能について

国立国会図書館電子情報部電子情報サービス課長  
竹内 秀樹

本日は、国立国会図書館の新しい検索システムである「国立国会図書館サーチ (NDL Search)」と、その中で実装を進めている「日中韓自動翻訳機能」の現状と課題について、報告します。このシステムは、2012 年 1 月の本格稼働を目指して、開発中のものです。昨年の 8 月に β 版(開発版)を公開し、機能追加を進めています。本格稼働まであと 2 か月ほどですが、現時点で、ほぼ全ての機能を公開しています。

国立国会図書館では、全国の公共図書館を対象とした総合目録データベースをはじめとして、各種の総合目録データベースを 1998 年から順次運用してきました。また、国内のデジタルアーカイブのポータルサイト(国立国会図書館デジタルアーカイブポータル)を 2007 年から運用しています。NDL Search は、これらのシステムを統合し一元化した新しい検索システムです。

### 1 NDL Search とは

NDL Search は、国内の各機関が持つ豊富な「知」を活用するためのアクセスポイントとなることを目指した検索サービスです。「国立国会図書館が保有しているか否かを問わず、冊子体に加えて、デジタル化された画像、テキスト、音声等の様々な形態の情報を、いつでも、どこでも、利用者が求める形で、迅速かつ的確に、アクセス又は案内できるようにすること」を目的としています。

これまでの OPAC は、図書館が所蔵する蔵書だけしか検索できません。また、紙資料を中心とした検索システムであるため、電子情報資源の検索ができないものがたくさんあります。検索対象が書誌情報だけで、資料の内容がわからない、一文字でも検索語の入力を誤るとヒットしないなど、インターネットの検索エンジンなどと比べて、機能性に劣る検索システムであるのが実情です。

これに対して、国立国会図書館の新しい検索システムは、国立国会図書館の蔵書はもちろんのこと、全国の公共図書館や大学図書館等で所蔵されている蔵書も統合検索することを可能とし、日本国内の知的情報資源の所在をワンストップで確認できるようにすることを目指しています。また、従来の図書館資料のほか、各種のデジタルアーカイブが提供する電子情報資源を含む、

様々な形態の情報資源を統合検索できるようにします。さらに、タイトルレベルに留まらず、雑誌については記事・論文レベルまで検索できます。資料の全文も検索の対象とすることができれば、書誌情報だけでは得られない、より掘り下げた検索が可能になります。現在、試行的に、国立国会図書館の刊行物を対象に全文検索を可能としています。今後、対象を拡大し、最終的には、利用者が必要とする事実や情報・知識そのものを検索できるようにすることを目指しています。

## 2 主要機能とシステム構成

### (1) 主要機能

NDL Search の主な機能は次のとおりです。

- 統合検索機能
  - 現在、65 のデータベースから収集した約 6 千万件の文献情報等を検索できます。このほかにも、メタデータのハーベストは行っていないものの、SRW/SRU、Open-Search 等のプロトコルにより、横断検索が可能なデータベースが、132 データベースあります。
  - 全文テキスト化された資料に関しては、書誌情報だけでなく、本文の検索ができます。
  - レファレンス事例などの事実情報も検索できます。
  - 検索結果を提示するだけでなく、可能な限り入手手段を案内します。国立国会図書館だけでなく、利用者の近在の図書館など、外部の資料所蔵機関を案内します。また、それらの機関で所蔵されていない場合には、Amazon 等のオンライン書店や、GoogleBookSearch といったオンラインサービスへもナビゲートします。
- 検索支援機能
  - 類義語・同義語による検索や、表記の相違を吸収した検索などが可能です。
  - 検索語を中国語、韓国語、英語に自動翻訳して検索できます。また、検索結果を自動翻訳して表示することができます。
- 検索結果のグルーピング機能
  - 検索結果を表示する際に、同一著作や関連する著作をグループ化して、FRBR 風に表示します。著作の関連性を視覚的にわかりやすく提示します。
  - 検索結果は、適合度順に排列できます。
- 検索結果の絞り込み機能と再検索機能
  - 資料種別、所蔵館等から絞り込み検索を行うことができます。(ファセット検索)
  - 関連キーワード等から再検索を行うことができます。(シソーラス検索、連想検索等)
  - このように、検索結果を効果的に絞り込んだり、広げたりすることが可能です。
- ブックマーク機能
  - 書誌情報の固定 URL 表示、Twitter への投稿機能、検索結果一覧の RSS 配信機能等、検索結果を活用するための様々な付帯機能を利用できます。

- 様々な用途に対応した GUI の提供
  - 障害者や高齢者等のユーザビリティ・アクセシビリティを考慮したデザイン・画面構成とし、点字図書や録音図書、大活字本などの障害者向け資料を検索する機能を提供しています。
  - 携帯電話・スマートフォンに対応した GUI を提供しています。
  - 成人だけでなく、子どもも検索サービスを利用できるように、子ども向けの GUI を開発しています。
- 外部サービス連携機能
  - このサービスを他のシステムから利用するための標準的な API を提供しています。まず、NDL Search の検索結果を、利用者自身の Web サービス上で利用できるようにする検索用 API を提供しています。また、メタデータをまとめて入手して利用できるようにするダウンロード用 API を提供しています。ダウンロード用 API では、新着図書情報の提供も行っています。新着図書情報とは、国立国会図書館に納入された国内刊行図書の基本書誌情報(作成中書誌)を、納入後数日で提供するサービスです。
  - APIを通じてデータを自由に利用できるようにすることによって、外部の機関がマッシュアップにより、新しい付加価値のある情報サービスを提供できるように支援します。

## (2) システム構成

NDL Search は、オープンソースソフトウェア (OSS) や外部のウェブサービスを積極的に採用して、開発することを基本方針としています。ベースとなるシステムには、日本で開発されたオープンソースの統合図書館システムである Next-L Enju を採用しています。そのほか、ウェブの収集ロボットである Heritrix、分散ファイルシステムである Hadoop、連想検索エンジンである GETAssoc、CMS である WordPress といった OSS を活用してシステムを構築しています。開発したソフトウェアも、将来的には OSS として公開し、公共図書館等での利用に供することを想定しています。

## 3 日中韓翻訳機能

### (1) 目的

NDL Search の検索対象は日本語の資料だけでなく、外国資料も含まれます。この中には、中国語、韓国語の文献も多数あります。こうした文献は、日本語ではなく、オリジナルの言語で検索しなければ、検索できません。検索語を日本語で入力した場合でも、他の言語に翻訳して検索し、その検索結果を日本語に翻訳できれば、様々な言語の資料の利用が容易になります。

### (2) 機能の概要

独立行政法人情報通信研究機構(NICT)が開発した「日中・中日言語処理技術の開発研究」

の実証実験システムを、WebAPI 経由で利用することとしました。これにより、検索キーワードを日本語から中国語に変換して、中国語文献等の検索を行う機能を実現しました。この実証実験システムは、中日の汎用的な用例翻訳手法及び科学技術文献の翻訳・情報検索用辞書を半自動作成する手法を利用したものです。ただし、現在の NICT の実証実験システムは、科学技術分野以外の領域には十分に対応できていません。このため、中国語、韓国語の翻訳サービスとして実績のある株式会社高電社の翻訳ソフトウェアを採用し、複数の翻訳サービスを自由に選択できるようにしました。

高電社は、機械翻訳ソフトウェアの開発に 20 年以上取り組んでいるベンダです。現在、韓国語・中国語・英語と日本語の双方向のテキスト翻訳機能を提供しています。中国語の翻訳エンジンは、清華大学の協力のもとで開発したものです。翻訳の基本アルゴリズムは、独自のルールベース翻訳方式を採用しています。加えて、翻訳原文を対訳文例データベースとマッチングさせ、ルールベースだけでは正確に翻訳できないフレーズの翻訳にも対応しているとのこと。基本辞書として、中日 30 万語、日中 38 万語を持ち、ユーザ辞書も追加可能です。中日翻訳で 80% 以上の認識率を持つということです。中国のオンラインショッピングサイトであるアリババ (Alibaba) でも、中国語翻訳エンジンとして採用されていると聞いています。

### (3) 翻訳サービスの流れ

翻訳サービスの流れのイメージは別紙スライドのとおりです。ここで、中日翻訳機能を、デモンストレーションします。(デモ操作)

#### (a) 検索トップページ

これがトップページです。単一の検索窓に検索語を入力する「簡易検索」と、複数の項目での検索が可能な「詳細検索」というモードがありますが、今回は時間の問題もありますので、簡易検索画面のみご紹介いたします。この中の「翻訳検索選択」をクリックすると、ご覧のとおり、「韓国語に翻訳」「中国語に翻訳」等、「(翻訳検索しない)を除き)5 種類の選択肢が示されます。本日は、このうち、「中国語に翻訳」を選択します。今回は、「中国」と「科学技術史」で検索を行います。これにより、両方の単語をメタデータに含む書誌を検索します。



< 検索トップページ >

## (b)検索結果一覧画面

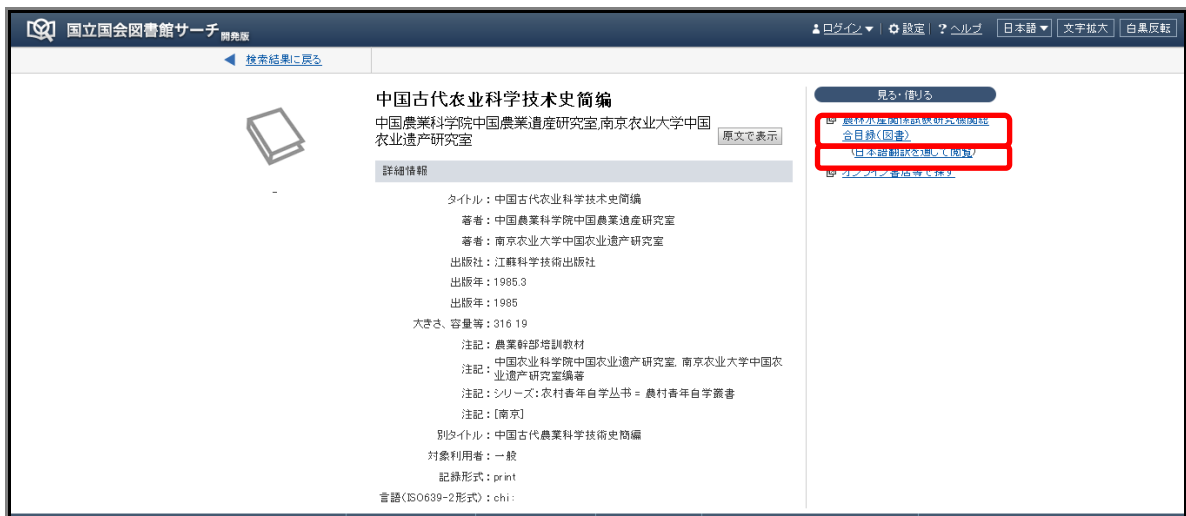
こちらが検索結果一覧画面です。入力したキーワード(原文)とそのキーワードの翻訳結果は、検索結果一覧画面の左上で確認可能です。真ん中の部分が、その「中国」「科学技术史」で検索を行った結果の一覧です。この状態では、中国語が読めない人にはどのような資料が分かりません。ここで、「日本語で表示」をクリックすると、一覧画面上で、タイトルが翻訳されます。この事例では、「中国の古代の農業の科学技術の史簡編」と表示されます。「原文で表示」をクリックすると元に戻ります。

The screenshot shows the search results page for the keyword '中国 科学技术史'. The page is in Japanese. At the top, there is a search bar with the keyword and a search button. Below the search bar, there are navigation tabs: 'すべて', '本・雑誌', '記事・論文', '新聞', '児童書', 'DVD・CD', 'デジタルアーカイブ', and '参考情報'. On the left side, there are filters for '検索結果の絞り込み', '資料種別', and 'データベース'. The main content area shows a list of search results. The first result is '中国 中国 科学技术史 科学技术史' with a '日本語で表示' button. The second result is '中国科学技术史 下巻' with a '日本語で表示' button. The third result is '中国科学技术史 上巻' with a '日本語で表示' button. The fourth result is '中国 古代农业科学技术史简编' with a '日本語で表示' button. The fifth result is '中国 中国科学院 中国农业遗产研究室 南京农业大学 中国农业遗产研究室 江苏科学技术出版社 1985' with a '日本語で表示' button. On the right side, there is a '関連キーワード' section with various related keywords and a '全て表示' button at the bottom right.

### < 検索結果一覧画面 >

## (c)書誌詳細画面

検索結果一覧画面でタイトルをクリックすると、その資料の詳細情報を見ることができます。ここでも「翻訳」ボタンを押すと日本語で詳細情報を見ることができるのですが、本日は省略させていただきます。画面の右側をご覧ください。ここに、提供元データベースへのリンクが表示されています。この資料は、「農林水産関係試験研究機関総合目録」（農林水産省所管の全国の試験研究機関が所蔵する図書資料類の総合目録）内のデータのようです。



< 書誌詳細画面 >

(d)提供元データベースの書誌詳細画面

書誌詳細画面で、提供元データベース名をクリックすると、原文のままの提供元データベースの書誌詳細画面が表示されます。次に、「日本語翻訳を通して閲覧」というリンクをクリックすると、日本語に翻訳された提供元データベースの書誌詳細画面が表示されます。



< 提供元データベースの書誌詳細画面 >



< 提供元データベースの書誌詳細画面(日本語翻訳) >

以上で、現時点で国立国会図書館サーチが実装している「日本語⇄中国語」翻訳機能のうち、「日本語で中国語資料を探す場合」のデモンストレーションを終了します。本日はご紹介しませんが、中国語で日本語資料を探すことも可能であり、また、英語、韓国語でも同様に翻訳機能を利用できます。

(4) 日中韓三国による検索、翻訳の実証実験

「日中韓電子図書館イニシアチブ協定」を昨年8月に締結しました。そこで掲げられた目標の実現策として、次の2点が挙げられています。第一に、NLC、NDL及びNLKは、メタデ

ータ・スキーマの標準化、ポータル統合及び電子情報への長期のアクセスを可能にするための共同開発を促進する。第二に、各図書館が有するそれぞれのポータルの相互運用性を高めることにより、三者の統合的情報サービスの第一段階を実現する。この合意に基づく最初の共同実験として、日本語、中国語、韓国語の言語の壁を超えた自動翻訳の実証実験を、日本から提案させていただいています。

NDL Search では、中国語、韓国語の資料の翻訳検索を行えるようにし、その有効性を検証しています。今後、双方向での検索・閲覧を実現することを視野に収めつつも、まずは、翻訳機能の実用性の検証のために、NLC、NLK の蔵書目録等を検索対象としてさせていただくことを提案しました。この結果、NLK との連携は、今年の 4 月に実現しました。中国国家図書館との間でも、担当者レベルで具体的な調整をさせていただいています。今後、作業を加速化したいと考えています。

次のステップとしては、中国、韓国で実験的に利用可能な翻訳システムがあれば同様に実装させていただくことが考えられます。また、翻訳精度を高めるために、三か国が協力して、用例集、用語の語彙の充実に取り組むことなども、考えられます。この成果を各国のポータルシステムに実装することにより、言語の壁を越えて、各国が保有している情報を相互に検索・閲覧できることを目指したいと思います。

#### 4 今後の課題

最後に、今後の課題について、簡単に触れます。利用者の情報探索の目的は、問題・課題の解決です。回答が掲載された資料の所在ではなく、回答そのものを知識・情報として得ることです。今後は、従来の単なる情報検索から、事実としての知識検索へ進化させ、知識の再利用による新たな知識の創造が求められています。それを実現するための前提として、資料をデジタル化し、内容を可視化して集積することが必要です。加えて、個別の情報に意味的にタグ付けし、知識として相互に関連付けて、利用者が求める知識として、よりの確に取り出せるようにすることが必要です。

「可視化」において、自動翻訳により言語差異を吸収し、また、交換されるメタデータの記述要素、記述規則を共通化し、類義語を把握するシソーラス、語彙の違いを吸収するオントロジー等の言語や表現の差異を吸収する技術を相互交換して適用することができれば、言語の壁を超えて、よりの確な検索が可能となります。これにより、「三国の統合的情報サービスの第一段階の実現」に近づけるものと考えています。

以上で報告を終わります。ありがとうございました。