

## 寄稿 4

# 中国・韓国語の特許文献を 日本語で検索可能なシステムのご紹介

櫻井 健太・立花 啓・上田 泰・磯部 洋一郎・金子 秀彦

### 抄録

特許庁は、中国・韓国語の特許文献の検索の利便性を向上させるために、「中韓文献翻訳・検索システム」の試行版の提供を平成26年11月13日に開始し、平成27年1月5日に本格稼働を開始しました。本システムは審査官のみならず、一般の方もご利用いただけます。本稿では、本システムをより多くの方にご利用いただくために、一般の方がご利用いただける機能を中心に紹介します。

## 1. 背景

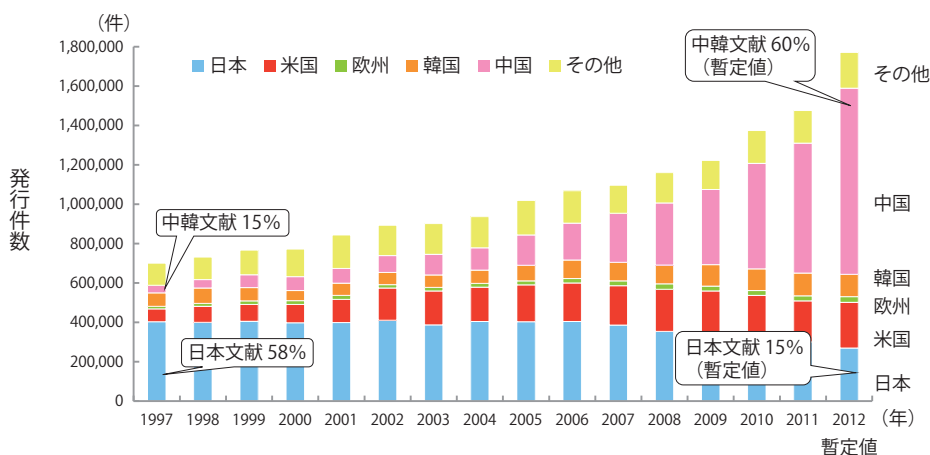
近年、世界の特許文献において、日本語以外の言語で記載された外国特許文献の割合が増加しています。特に、中国・韓国語の特許文献については、出願件数の急激な増加により、中国・韓国語でしか読むことのできない特許文献が、世界の特許文献の約6割を占めるに至っています。このような状況の中、我が国企業等の研究開発を推進するため、これらの特許文献を検索できる環境の整備が求められています。

## 2. 中韓文献翻訳・検索システムの概要

このような状況に対応するため、中国・韓国語の特許文

献の検索環境の整備を最優先で進めています。そこで、「中韓文献翻訳・検索システム」の試行版の提供を平成26年11月13日から開始し、平成27年1月5日から本格稼働を開始しました。本システムの主な特徴は以下のとおりです。

- 中国・韓国語の特許文献を日本語に機械翻訳して蓄積することにより、日本語でこれらの文献の全文を検索することが可能です。機械翻訳には、特許庁が作成した約100万語の中日辞書を使用しており、翻訳精度の向上を図っています。詳細については、4.で後述します。
- 本システムはクラウド上に構築されています。インターネットを介して、審査官のみならず、一般利用者也無料でご利用いただけます。本稿末尾に記載のURLからアクセスいただけます。



(備考) 世界で発行された特許文献を言語別に整理し、重複を排除したものの。日本語、米国(英語)、欧州(英語、仏語、独語)、韓国(韓国語)、中国(中国語)の順で該当する国・地域(言語)の特許文献としてカウント。

図1 世界の特許文献

○照会・検索の対象となる文献は、中国特許出願公開公報、中国特許公報、中国実用新案登録公報、韓国特許出願公開公報、韓国特許公報、韓国実用新案公開公報、韓国実用新案登録公報です。

○本格稼働を開始した時点で、以下の範囲の文献が蓄積されています(合計約1000万件)。本格稼働後も、平成26年8月以降に発行された文献を順次蓄積し、中国・韓国における文献の公開から翻訳文の照会・検索が可能となるまでの期間を短縮していく予定です。

中国文献：平成15年～平成26年7月公開分

韓国文献：平成15年～平成26年7月公開分

### 3. 中韓文献翻訳・検索システムの機能

#### (1) 公報テキスト検索

図2の検索画面で検索を行います。検索項目(要約、発明の名称、公報全文等)を選択し、検索キーワードを入力するだけで検索式を作成できるので、特許検索に詳しく無い方でも比較的簡単に検索を行うことができます。

また、以下のような高度なテキスト検索機能もご利用いただけます。

#### ○NOT検索機能

除外したい検索キーワードを指定することで、指定した検索キーワードを含まない文献を検索することができます。

#### ○近傍検索機能

複数の検索キーワードとキーワード間の文字数を指定することで、当該複数の検索キーワード同士が、指定した文字数

以内に近接して出現する文献を検索することができます。

#### ○拡張NOT検索機能

検索キーワードと除外検索文字列を指定することで、当該検索キーワードを含む文献集合から、除外検索文字列のみを含む文献を除外して検索することができます。例えば、検索キーワードとして「ログ」、除外検索文字列として「プログラム」を指定した場合、文献中に「ログ」を含む単語が「プログラム」だけである場合はヒットせず、文献中に「ログ」を含む単語が「アナログ」と「プログラム」の二つある場合はヒットします。

#### ○検索結果集合式機能

検索結果集合式欄の左側のチェックボックスにチェックを入れることで、検索結果エリアに表示されている文献集合と、新たに指定した検索条件の文献集合との積集合(AND)、和集合(OR)を作成することができます。

#### (2) 公報番号索引照会

図3の公報番号索引照会画面で、個別に文献番号を入力して番号照会を行い、発行国、発明の名称や出願人名等を一覧表示することができます。また、番号照会を行った文献の翻訳文を表示することができます。

#### (3) 翻訳文表示

図4の翻訳文表示画面で、上記公報テキスト検索及び公報番号索引照会の結果得られた文献の内容を表示します。左側に機械翻訳文、右側に図面が表示されます。機械翻訳文の表示では、検索式で指定したキーワードがハイライトされます。

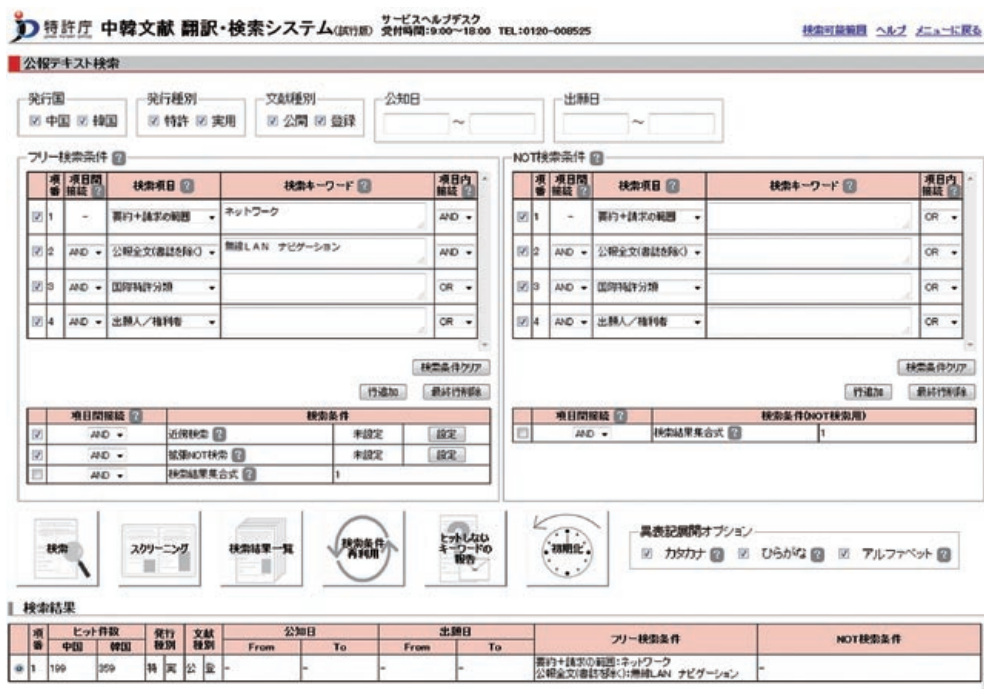


図2 検索画面



図3 公開番号索引照会画面

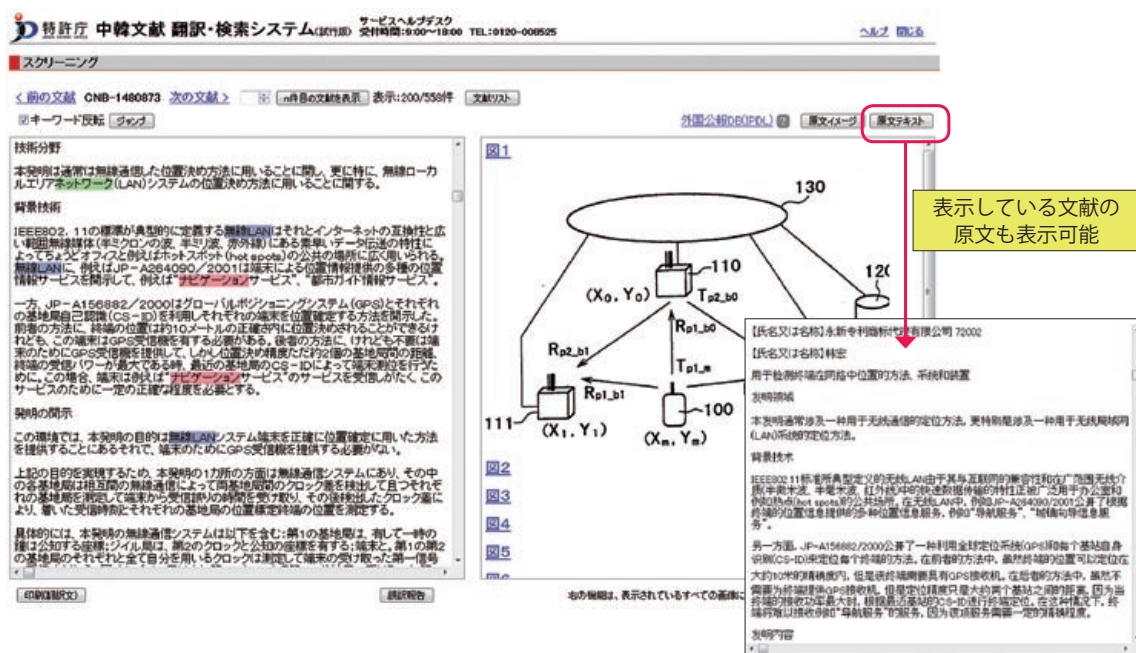


図4 翻訳文表示画面

また、「原文テキスト」ボタンを押下することで表示している文献の原文を別画面で表示することも可能です。

#### 4. 翻訳品質の向上

翻訳品質をいかに向上させるかが本システムの重要な課題の一つです。本システムでは、翻訳品質を向上させるために、以下のような取組みを行っています。

##### (1) 機械翻訳用対訳辞書の作成

平成23年度に特許庁が行った調査によれば、民間事業者が提供する日英機械翻訳ソフトウェアにおいては通常100万語以上の辞書が備えられているのに対して、中日機械翻訳ソフトウェアの辞書の基本語は30万語程度であり、十分な翻訳精度の確保のためには、中日辞書を新たに作成

する必要があることが分かりました。

この調査の結果を受け、特許庁では、平成24年度以降毎年度、特許文献の翻訳に特化した専門用語中日辞書を作成する事業を実施しています。具体的には、中国語と日本語の対(対訳コーパス)を作成し、さらに、対訳コーパスを単語に分割して単語単位での対(対訳辞書候補)を作成するところまでを機械的な処理によって行い、最後に、対訳辞書候補を手手でチェックして誤っている語などを削除することで、辞書を作成しています。また、平成25年度からは、中国語と日本語の対訳コーパスに加え、中国公開特許の要約とそれを人手で日本語に翻訳した和文抄録の対も辞書作成の材料として使用しています。

このような辞書作成事業によって、平成24年度には、約100万語の中日辞書を作成しました。本システムに蓄積

した文献の翻訳には、市販の辞書に加えて、本辞書も利用しています。また、平成25年度以降に作成した辞書についても今後蓄積される文献の翻訳に利用する予定です。

## (2) 誤訳報告

本システムの翻訳文に誤訳を発見した場合、図5の誤訳報告画面から誤訳報告を行うことができます。利用者からの誤訳報告を中日辞書・韓日辞書に反映させることで更なる翻訳品質の向上を図る予定ですので、誤訳報告にご協力いただけますようお願いいたします。

既に翻訳された中国文献・韓国文献についても、翻訳品質の向上を図るため、辞書の追加(上記(1))及び利用者からの誤訳報告(上記(2))等を反映した最新の翻訳エンジンを利用した再翻訳作業の実施を今後予定しています。

## 5. おわりに

本稿では、中韓文献翻訳・検索システムの機能について、一般の方がご利用いただける機能を中心にご紹介しました。本システムは、出願時の先行技術文献調査、中国・韓国における権利調査等の際のご利用が想定され、従来の文献調査業務の大幅な効率化が見込まれます。是非ご活用下さい。

〈中韓文献翻訳・検索システム〉

<http://www.ckgs.jpo.go.jp>

(サービス提供時間：開庁日の8時から22時まで)

## Profile

櫻井 健太 (さくらい けんた)

平成15年4月 特許庁入庁(審査第一部計測)  
平成21年4月 経済産業省商務情報政策局  
平成24年7月 スタンフォード大学留学  
平成26年1月 総務課情報技術統括室

立花 啓 (たちばな あきら)

平成19年4月 特許庁入庁(審査第二部一般機械)  
平成26年1月 総務課情報技術統括室  
平成27年1月 審査第二部動力機械

上田 泰 (うえだ やすし)

平成17年4月 特許庁入庁(審査第一部自然資源)  
平成22年10月 審査第一部審査調査室  
平成23年10月 審査第一部材料分析  
平成25年4月 総務課情報技術統括室

磯部 洋一郎 (いそべ よういちろう)

平成20年4月 特許庁入庁(審査第三部医療)  
平成24年4月 審査第三部有機化学  
平成25年7月 総務課情報技術統括室  
平成27年1月 審査第三部医療

金子 秀彦 (かねこ ひでひこ)

平成17年4月 特許庁入庁(審査第四部情報記録)  
平成22年1月 調整課審査企画室  
平成23年1月 審査第四部映像システム  
平成23年7月 米国カーネギーメロン大学に留学  
平成24年7月 審査第四部情報処理  
平成25年7月 総務課情報技術統括室  
平成27年1月 審査第四部伝送システム

図5 誤訳報告画面