

中韓文献翻訳・検索システム

Machine Translation and Retrieval System of Chinese and Korean Documents

特許庁 総務部総務課特許情報企画室長 **檀本 英吾**

PROFILE

1996年4月特許庁入庁、光デバイス等の特許審査・審判に従事。他、審判課審判企画室（課長補佐）、総務課情報技術企画室（情報技術企画班長）等を経て、2014年4月から現職

1 はじめに

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急増している。特に中国特許文献及び韓国特許文献（中韓文献）の増加が著しい。1997年には世界の特許文献の15%であった中韓文献が、2012年では60%を占めている¹。中韓文献を審査官及び外部ユーザーが容易に調査できる環境を実現しなければ、不十分な先行技術文献調査による権利の安定性の低下に加え、日本企業が進出先において現地企業から訴えられる可能性の増大等のリスクが高まるおそれがあることから、機械翻訳等を活用した日本語による中韓文献の検索環境の整備が喫緊の課題である。

この課題に対応すべく、特許庁は、中韓文献の機械翻訳による日本語全文テキストを作成し、日本語による全文テキスト検索を可能にするシステムを2015年1月にリリースする予定である。

このシステムのサービスや機能について、以下に紹介する。なお、現在、設計・開発中であるため、変更があり得る。

2 サービス概要

(1) 機械翻訳

中国特許出願公開公報、中国特許公報、中国実用新案公報、韓国特許出願公開公報、韓国特許公報、韓国実用新案公報を日本語翻訳し、蓄積する。つまり、文献の照会時に随時翻訳するのではなく、予め翻訳をして蓄積する。いずれの公報も2003年以降発行されたものが対象となります。蓄積文献数は、リリース時が約800万件、その後毎年200万件程度の文献を追加する見込みである。

(2) 検索

検索機能としては、発行国・発行種別・文献種別の指定、公知日などの日付範囲検索、翻訳文のフルテキスト検索、IPC検索等が可能となっている。

また、文献番号による文献番号索引照会も可能である。

(3) 表示

翻訳文の表示、原文テキスト表示が可能である。原文イメージを見るためにEspacenetへのリンクボタンも設けられている。

翻訳文の表示では、フルテキスト検索で使用されたキーワードがハイライト色でキーワードハイライトされる。

1 世界で発行された特許文献（実用新案も含む）を言語別に整理し、重複を排除したもの。複数の国に出願され、公開された同内容の特許文献について、日本語があるものは日本の特許としてカウント。日本語がない場合は、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。

3 サービス詳細

(1) 機械翻訳

機械翻訳では、新しく出てきた専門用語等への対応を実施することとなっている。

具体的には、機械翻訳する際に機械翻訳辞書に登録されていないと検出された未知語情報や、利用者及び誤訳パトロールにより報告された誤訳等の情報に基づいて、辞書の改版や翻訳メモリ等への追加登録、翻訳エンジンのパラメータ調整等を行う。これにより辞書及び翻訳エンジンが改善され、新しく出てきた専門用語等にも対応することができるようになる。

また、中日の機械翻訳辞書は英日の辞書ほど利用できるものが多くないため、中日の Patent Family 等から対応する中日の単語対を機械的に抽出し、それを人手

で確認することによって特許文献に特化した辞書を新たに作成し、これを本サービスの機械翻訳辞書として使用している。

作成された翻訳文は、以下の検索・照会によるサービス提供の他、バルクのデータとしての提供も予定されている。

(2) 検索

図 1 の検索画面で検索式を作成する。

検索式作成は、詳しく無い方でも比較的簡単に検索できるように、検索項目を選択し検索キーワードを入力することで作成できるようになっている。図 1 では、「フリー検索条件」に 4 つの項目（4 行）が表示されているが、行を追加することができる。NOT 検索は別途設けた「NOT 検索条件」に入力することで実行できるため、複雑な検索式を作成し易くなっている。

The screenshot shows the 'Patent Text Search' interface. At the top, there are filters for 'Country of Issue' (China, Korea), 'Document Type' (Special, Practical), and 'Document Category' (Public, Registered). Below these are two main search condition areas: 'Free Search Conditions' and 'NOT Search Conditions'. Each area contains a table with columns for 'Item No.', 'Item Category', 'Search Item', 'Search Keyword', and 'Item Internal Category'. The 'Free Search Conditions' table has 4 rows with 'AND' logic between them. Below these are buttons for 'Clear Search Conditions', 'Add Row', and 'Delete Last Row'. At the bottom of the search area are icons for 'Search', 'Screening', 'View Search Results', 'Use Search Conditions', 'Report (Hit/Off)', and 'Refresh'. To the right, there are options for 'Display Table Expansion Options' including 'Katakana', 'Hiragana', and 'Alphabet'. At the bottom, a 'Search Results' table is shown with columns for 'Item No.', 'Hit Count', 'Country', 'Document Type', 'Document Category', 'Publication Date', 'Issue Date', 'Search Conditions', and 'NOT Search Conditions'. The first row shows 6 hits for item 1.

図 1 検索画面

その他、拡張 NOT 検索、近傍検索、異表記展開（カタカナ、ひらがな、アルファベット）が可能となっている。拡張 NOT 検索とは、例えば、検索条件「ログ」、否定条件「プログラム」で検索すると、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2つがある場合はヒットする。

それぞれの機能について不明な点があれば、アイコンや「？」マークにマウスオーバーすることで、説明文が表示される。

(3) 表示

照会画面では、左側に翻訳文、右側に図面が表示される。

特許庁 中韓文献 翻訳・検索システム ヘルプ 閉じる

スクリーニング

< 前の文献 CNB-1480873 次の文献 > 何回目の文献を表示 文献リスト 表示: 1/1件

キーワード反転

正確に端末ネットワークにある位置を検出しGPS受信機の方法、システムと装置を必要としないことに用いることを提供した。一例に、位置判断端末に用いる方法は基地局を計算する位置を含む；基地局と他方の基地局間のクロック差を検出する；基地局を測定して端末から信号の受信時刻を受け取る；少なくとも基地局の位置を利用すること、クロック差と、受信時刻位置判断端末と。

【図1】

印刷/翻訳文) ?

図1

図2

右の機能は、表示されているすべての画像に対して有効になります。

図2 翻訳文表示画面

翻訳文の表示では、検索式で指定した語句が検索項目毎に色を変えてキーワードハイライトされる。また、画面は拡大縮小表示が可能となっている。

右上の「原文イメージ」ボタンを押下すると、Espacenetにより原文イメージを照会できる。また、「原文テキスト」ボタンを押下すると、ロボットアクセス防止のための認証機能を経た後、以下のように原文のテキストデータを照会できる。

4 おわりに

機械翻訳の技術は発展途上であり、機械翻訳文から原文の文意がただちに分からない場合も想定される。特に中国文献の日本語への翻訳は、英語や韓国語から日本語に比べて精度が高くない。しかし、中国語又は韓国語を読める審査官及びユーザーは少ないことから、本サービスにより中韓文献へのアクセス性が向上することが期待される。



図3 原文表示画面