

高精度でセキュアな英文特許自動翻訳の提供開始

【ポイント】

- 大規模な対訳データを活用した自動翻訳エンジンにより、従来よりも大幅に翻訳精度が向上
- 特許翻訳固有の問題解決で化合物表記・図・表を含む特許公報の表記通りの正確な翻訳が可能に
- 企業の情報機密管理上重要な特許情報をセキュアなクラウド上で、大規模に翻訳するサービスを展開

国立研究開発法人情報通信研究機構(NICT、本部: 東京都小金井市、理事長代行: 黒瀬 泰平)、株式会社みらい翻訳(みらい翻訳、本社: 東京都渋谷区、代表取締役社長: 栄藤 稔)、日本特許翻訳株式会社(NPAT、本社: 東京都中央区、代表取締役社長: 本間 奨)、一般社団法人化学情報協会(JAICI、東京都文京区、理事長: 小沼 和彦)は、高精度でセキュアな英文特許自動翻訳サービスを開発しました。

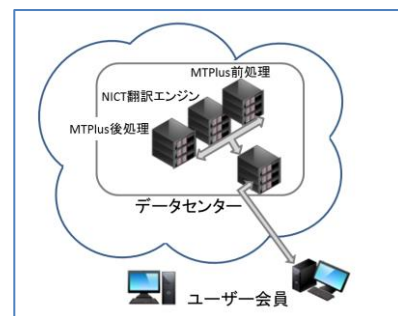
本特許自動翻訳サービスは、NICT が開発した高精度な英日特許自動翻訳エンジンに対して、NPAT と JAICI が読みやすさと正確性を向上させるための技術を適用し、みらい翻訳がセキュアなクラウド環境で提供するものです。本サービスは、2017年4月から提供を開始する予定です。

【背景】

近年、海外で増加している特許の侵害・訴訟リスクを回避するため、外国の特許文献の検索や内容把握が不可欠となっており、日本語への自動翻訳ニーズが高まっています。特許翻訳では、大量の専門用語の訳語を正確に選択することが求められます。加えて、翻訳対象の特許原文は、他者に漏えいしないように厳重に守る必要があります。

【今回の成果】

- NICT は、従来の NICT 英日特許自動翻訳エンジンのアルゴリズムを改良すると同時に、大規模な対訳データを構築することにより、大幅に精度向上した自動翻訳エンジンを開発しました。
- みらい翻訳は、NICT の英日特許自動翻訳エンジンに対して、速度・可用性の向上を達成し、商用ライセンスを他者に提供するとともに、企業の情報機密管理上重要な特許情報をセキュアなクラウド上で、大規模に翻訳するサービスを展開します。
- NPAT と JAICI は、これまで翻訳することが難しかった特許文書固有の問題を解決しました。特に、医薬・化学分野での文献や物質検索、特許調査等に必須となる化合物表記の厳密な翻訳が可能になりました。



英文特許自動翻訳サービスのイメージ図

【今後の展望】

今回開発した「英日特許自動翻訳ソフトウェア」によって、みらい翻訳と NPAT は、2017年4月から、米国特許公報、欧州特許公報、国際出願公報の英日自動翻訳サービスを開始します。

NICT は、特許を含む多分野・多言語での長文翻訳の品質を今後更に改善するために、新たな技術の研究を推進してまいります。

【今回の成果の詳細】

NICT は、特許などの長文の自動翻訳技術の高精度化を目指して自動翻訳技術の高度化に取り組んでおり、2013年3月には英語・中国語から日本語への特許自動翻訳を商用ライセンスし、2014年7月、2016年4月には特許庁と多言語特許文献の高精度自動翻訳の実現に向けて協力合意をしました^(注1)。さらに、これらの実用化が評価され2014年6月には、第9回AAMT長尾賞を受賞^(注2)しました。

今回、NICT は、従来のNICT英日特許自動翻訳エンジンよりも精度が大幅に向上した自動翻訳エンジンを開発しました。その精度向上の理由の一つは、特許庁との協力合意に基づく研究成果を活用して、日本特許と英語で出願された特許(米国特許、欧州特許など)のpatentファミリーから、従来比10倍超となる3.5億文の日英特許対訳データを構築し、エンジン開発に活用したことです。また、第9回AAMT長尾賞を受賞し、既に高精度であることが確認されている「語順変換と訳語選択の独立実行による統計翻訳技術」を更に改良することにより、コアとなる英日特許自動翻訳エンジンを構築しました。

みらい翻訳は、NICTの英日特許自動翻訳エンジンに対して、速度・可用性の向上を達成するとともに、セキュアなクラウド上で、大規模にサービス展開します。

NPATは、みらい翻訳が提供するNICTの英日特許自動翻訳エンジンに対して、特許特有の文体を解析して、翻訳精度と読みやすさを向上するMT Plus技術^(注3)を適用することにより、読みやすく、かつ、図・表・数式を含めた特許公報の表記通りの正確な日本語訳を提供します。

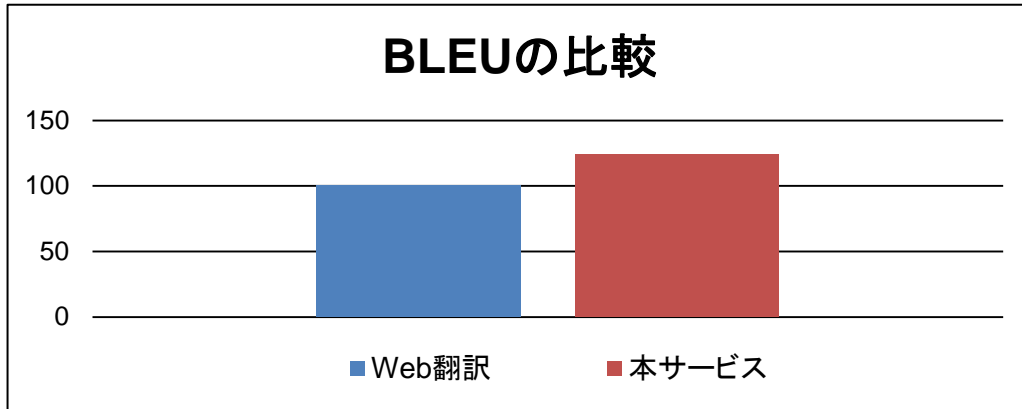
本サービスの特徴は、次の4点です。

- ① 高精度なNICT英日特許自動翻訳エンジンに対して、MT Plus技術を用いることにより、特許特有の非常に長い文章に対応可能となりました。この結果、「訳抜け」や「請求項において、主題に構成要素の記述が混入することや、構成要素間の記述が前後すること」を解消しました。
- ② これまでの統計的自動翻訳エンジンでは正確に翻訳することが難しかった重要低頻度語の問題を改善しました。具体的には、JALCIの化合物表記翻訳システムとMT Plusを連携させ、主に医薬又は化学分野での文献や物質検索、特許調査等に必須となる化合物表記の厳密な翻訳を可能にしました。この結果、化学情報の正確な翻訳が必要なユーザーにも、本サービスを利用していただけることが期待されます。
- ③ 諸外国(米国特許商標庁、欧州特許庁など)の外国公報XMLに準拠した成果として、テキストだけでなく、表、数式や図を含めた、外国公報全体を正確に翻訳できるようになりました。
- ④ 従来のクラウドサービスでは、意図しない情報漏えいの問題が再三指摘されており、IPA(独立行政法人情報処理推進機構)から注意喚起がなされています^(注4)。特に、企業の重要機密情報である特許調査・特許出願に関する情報はセキュアであることが極めて重要です。NICT翻訳エンジンによる本翻訳サービスは、フロア全体が自社データセンターとなっているデータセンター内に設置されたみらい翻訳がサポートする翻訳サーバーを用いて行われ、会員限定のサイトなどによる情報授受が閉じた中で行われるため安全なサービスとなっています。

具体的には、本サービスでは、会員から特許公報番号をいただき、会員メールアドレスあてに特許公報の翻訳結果のダウンロードリンクを提供します。セキュリティではファイアウォールを設備し、メールによる相手の確認を行います。

【翻訳精度の向上】

次ページの図に示すように、本サービスにおける英日特許自動翻訳の精度(本サービス)を、Web上にパブリックで利用可能な自動翻訳サービス(Web翻訳)と比較しました。翻訳システム間の精度の傾向を把握するため、BLEU^(注5)という自動翻訳の精度を比較するための一般的な尺度を用いて、定量的に比較しました。この図は、様々な外国公報の請求項や詳細な説明から抜粋した文(長文も短文も混在)を原文とし、対応日本特許又は人手翻訳を正解として、合計567文からBLEUを計算したものです。比較のために、Web翻訳のBLEU値を100%として、それとの比率を表示しています。MT Plus技術を適用した本サービスは120%を示しており、Web翻訳よりも高精度であることがわかります。BLEUは自動評価値ですが、20%の精度向上は、低頻度重要単語の精度向上など、体感的にも十分実感できるものです。



本特許自動翻訳サービスとWeb 翻訳の比較

<役割分担>

NICT: コアとなる高精度英日特許自動翻訳エンジンを開発

みらい翻訳: セキュアなクラウド上で4言語に対応した特許自動翻訳エンジンプラットフォームサービスを提供

NPAT: NICT英日特許自動翻訳エンジンにMT Plus技術を適用して高精度化

JAICI: 化合物表記翻訳システムを開発

<用語解説>

(注1) NICTの自動翻訳技術についての報道発表

○2013年3月21日、「英語特許文」の高精度「自動翻訳ソフトウェア」を開発

<http://www.nict.go.jp/press/2013/03/21-1.html>

○2013年3月28日、NICTの高精度な中日自動翻訳ソフトウェアがJapioのサービスに

<http://www.nict.go.jp/press/2013/03/28-1.html>

○2014年7月28日、NICTと特許庁が多言語特許文献の高精度自動翻訳の実現に向けて協力合意

<http://www.nict.go.jp/press/2014/07/28-1.html>

○2016年4月1日、NICTと特許庁の特許文献の機械翻訳に関する協力の継続について

<http://www.nict.go.jp/info/topics/2016/04/160401-1.html>

(注2) 第9回AAMT長尾賞を受賞

2014年6月19日、第9回AAMT長尾賞を受賞 <http://www.nict.go.jp/info/topics/2014/06/140619-2.html>

(注3) MT Plus技術

NPAT 開発の技術で NICT 自動翻訳エンジンの前処理・後処理をする。特許特有の長文を、主題と構成要素を特定してから自動翻訳することで、訳語の適切性を維持しつつ特許文の語順に適合することが可能。このために、前処理で英文手がかり句分析を行い、前処理文を翻訳した後、後処理で復元処理を行う。翻訳システムとしては複雑になるが、翻訳システムの制限による訳抜けを防止するとともに高精度で読みやすい文章が生成される。

特許請求項例文)

A system according to claim 4, the controller including: a) A control store for storing location data, the location data indicating a location for each display at the remote site; and, b) A control processor, the control processor being adapted to: i) Determine the information type of the respective information to be displayed; ii) Determine the location in which the respective information is to be displayed, the location being determined in accordance with the information type; iii) Schedule the display of the respective information at the respective location; and, iv) Transfer the corresponding display data to the display system.

NICT+MT Plus)

制御装置であって、a) 遠隔地における各表示のための位置を表す位置データを記憶するための制御および、b) i) 表示されるべき各情報の情報種別を決定し、ii) それぞれの情報位置が表示されるべきかを決定するため、位置は情報の種類に応じて決定され、iii) それぞれの位置で各情報のディスプレイを予定することおよび、iv) 表示装置に対応する表示データを転送すること、に適応された制御プロセッサを含む、請求項4のシステム。

また、統計的自動翻訳の弱点である低頻度語、例えば

5-(3'-(3''-(2,4-diamino-6-ethylpyrimidin-5-yloxy)propoxy)phenyl)-6-ethylpyrimidine-2,4-diamine

のような化合物表記も JAICI 化合物翻訳システムを前処理に組み込むことで

5-(3'-(3''-(2,4-ジアミノ-6-エチルピリミジン-5-イルオキシ)プロポキシ)フェニル)-6-エチルピリミジン-2,4-ジアミン

のように正しい翻訳を可能としている。

(注4) 情報漏えいに関する注意喚起

2015年2月20日、【注意喚起】クラウドサービスに入力した内容の意図しない情報漏えいにご注意

<http://www.ipa.go.jp/about/press/20150220.html>

(注5) BLEU

自動翻訳の性能を評価する自動尺度の一つ。BLEU では、自動翻訳の結果の訳文と、参照訳としての人手による訳文との類似度が高いほど、高い数値となる。

< 本件に関するお問い合わせ >

国立研究開発法人情報通信研究機構
先進的音声翻訳研究開発推進センター
内山 将夫
Tel: 0774-98-6343
E-mail: ltg-info@nict.go.jp
<http://www.nict.go.jp>

株式会社みらい翻訳
セールス&マーケティング部
藤原 祥造
Tel: 03-6434-1340
E-mail: info@miraitranslate.com
<https://miraitranslate.com>

日本特許翻訳株式会社
営業
鈴子 真由美
Tel: 03-5652-8935
E-mail: m-search@npat.co.jp
<http://www.npat.co.jp>

一般社団法人化学情報協会
情報技術部
一ノ瀬 桂子、山口 政隆
Tel: 03-5978-3606
E-mail: kichinose@jaici.or.jp,
myamaguchi@jaici.or.jp
<https://www.jaici.or.jp>

< 広報 >

国立研究開発法人情報通信研究機構
広報部 報道室
廣田 幸子
Tel: 042-327-6923
E-mail: publicity@nict.go.jp

株式会社みらい翻訳
セールス&マーケティング部
藤原 祥造
Tel: 03-6434-1340
E-mail: info@miraitranslate.com

日本特許翻訳株式会社
営業
鈴子 真由美
Tel: 03-5652-8935
E-mail: m-search@npat.co.jp

一般社団法人化学情報協会
広報担当
Tel: 03-5978-3608
E-mail: koho@jaici.or.jp