

多言語翻訳プラットフォームによる翻訳サービスの実装

長瀬友樹、大倉清司、富士秀、徐国偉、潮田明

(株)富士通研究所

nagase.tomoki@jp.fujitsu.com

1. はじめに

近年企業活動のグローバル化が進み、社内コミュニケーションや情報共有などで機械翻訳を利用する場面が増えている。機械翻訳は品質も向上しサーバ型製品でも手頃な価格で購入できるようになった。しかしながら、製品によって使い勝手は様々であり、例えば3ヶ国語以上を扱う部門では翻訳方向によって複数の製品を使い分けなければならないという問題があった。これは人間が翻訳ソフトを使う場合にもシステムが翻訳機能を関数として呼出す場合にも言える共通の課題である。

そこで、我々は既存のあらゆる翻訳エンジンを共通のインターフェースで利用することができる枠組みを考案し、その枠組みに沿って実際にシステムの設計と構築を行った（以降この枠組みのことを「多言語翻訳プラットフォーム」と呼ぶ）。

本稿では、まず多言語翻訳プラットフォームの概要とシステム構成を示し、その後、多言語翻訳プラットフォームの各機能と利用者のメリットについて説明する。最後に、多言語翻訳プラットフォームを利用して作成した検証用システムについて紹介を行う。

2. 多言語翻訳プラットフォームの設計

多言語翻訳プラットフォームを利用したシステム構成を図1に示す。システムは3階層になっており、第1階層が翻訳エンジンを呼出すアプリケーション、第2階層が翻訳プラットフォームのメインプロセス、そして第3階層が各ベンダによって提供される翻訳エンジンである。第2、第3階層を総称して多言語翻訳プラットフォームと呼ぶ。メインプロセス（第2階層）と翻訳エンジン（第3階層）の間のインターフェースを標準化し、その標準インターフェースに繋がる口を各翻訳エンジン側に用意することで、任意の翻訳エンジンを多言語翻訳プラットフォームに追加することが

きる。さらに、各種アプリ（第1階層）とメインプロセス（第2階層）の間のインターフェースを標準化することで、利用者は翻訳プラットフォームに組み込まれたすべての翻訳エンジン（第3階層）を共通の手順、共通のインターフェースで利用できるようになる。

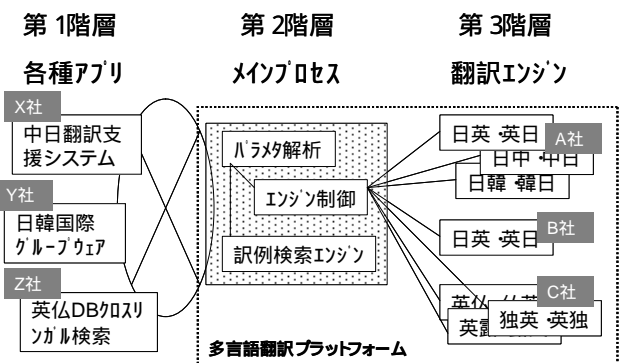


図1 システム構成

2.1 実装機能

翻訳エンジンは言語処理の集積体であり、ほとんどの翻訳エンジンは形態素解析、構文解析、辞書検索、文生成といった要素機能をプログラムの独立した形で内包している。利用者アプリ側から要求される機能という観点では、翻訳機能以外に単語分割や辞書引きといった要素機能を単独で利用されるケースも多いため、翻訳プラットフォームの実装機能として以下の4機能を備えることにした。

- 翻訳
- 単語分割
- 辞書検索
- 訳例検索

要素機能（単語分割、辞書検索）を実装するもう一つの理由は、翻訳エンジンがダイレクトに翻訳することができる言語対が未だ少ないという事

実である。つまり例えば日本語-タイ語のように翻訳エンジンはなくても辞書検索エンジンなら存在するという言語対は、辞書検索機能だけでも翻訳プラットフォームの機能として追加されることを狙っている。

訳例検索は翻訳エンジンの要素機能ではないが、産業翻訳への適用では必須の機能であることを鑑み4つ目の実装機能とした。訳例検索機能は言語非依存の共通モジュールとしてメインプロセス（第2階層）に組込まれている。

2.2 呼出し手順

利用者アプリから多言語翻訳プラットフォームの呼出し手順の設計は、呼出し元の OS、言語に依存しないこと、インプリメントが簡単なことを条件として検討を行った。通信手順の候補としては、DCOM[3]、CORBA[1]、SOAP[2]などを検討したが、DCOM、CORBAはいずれも OS 依存性が残り手順として適当でないことが判り、結果として SOAP を選択した。SOAP の採用によって上記の条件を満たすほか、以下のような利点が生まれる。

- ・ HTTP で通信を行うためファイアウォール越えが可能（社内 LAN からインターネット経由での遠隔利用が可能）
- ・ 近年普及し始めている Web サービスとして提供が可能

2.3 呼出しメソッド（インターフェース）

各実装機能の呼出しメソッドは基本的に以下をパラメータとした SOAP 呼出しメソッドである。

- *原文
- *翻訳方向（単語分割は処理言語）
- ベンダー
- エンジン識別子
- 分野
- 専門辞書
- オプション

ベンダー以下のパラメータは省略が可能であり、省略された場合はメインプロセスが保持しているデフォルト値によって処理される。「エンジン識別子」は同一ベンダーの中でエンジンを特定するために使用する。

処理結果はメソッドの復帰値として得られる。例えば翻訳機能の処理結果（XML）には以下の情報が含まれる。

- 翻訳結果
- 形態素解析結果（原文、訳文）
- 原文、訳文の単語対応情報

- 訳語候補単語

処理結果として翻訳結果のテキストのみを返すシンプルメソッドを別に持っており、詳細情報を返すことができないエンジンについては、シンプルメソッドのみをサポートすることもできる。

3. 多言語翻訳プラットフォームの利点

エンジン接続のためのインターフェース仕様と利用のためのメソッドが公開され Web サービスとしての多言語翻訳プラットフォームがフィールドで認知されるようになれば、世界中からネットワーク経由の機械翻訳利用が活性化されて機械翻訳の市場が広がり、それがさらに新しい開発の動機付けになるという好循環が生まれることが期待される。その根拠を3つの側面から説明する。

3.1 翻訳エンジンベンダーの利点

現在、商用システムとして市販されている大部分の機械翻訳製品は、資金力が豊富なメーカの製品である。機械翻訳の製品化は翻訳エンジンの他に GUI 開発やプロモーションに膨大なコストがかかるため、小規模な企業ではリスクが高く新規参入が難しい状況である。同じ理由で大手メーカであっても希少言語対の翻訳ソフトの製品化は困難である。

多言語翻訳プラットフォームによって、翻訳エンジンベンダはエンジンさえ開発できればほとんどコストをかけずにユーザを獲得することができるため、多言語翻訳プラットフォームの利用を前提としたベンチャー企業や大学による希少言語の翻訳エンジン開発に拍車がかかるものと期待する。

3.2 システム開発部門の利点

企業内システム開発部門またはインターネットポータルサービスの開発部門が言語処理機能を使う場合、従来はサーバ製品を購入して自社内でサーバを運用する必要があった。多言語翻訳プラットフォームを使えば、部門のアプリケーションから Web サービスを呼出すだけで済み、製品購入もサーバ運用も必要なくなる。さらに、プログラムコードを修正することなく対象言語対の増加に対応できる。多言語翻訳プラットフォームには翻訳をはじめ言語処理機能が集約されているので、迅速なトライ&エラーにより最適なエンジンの組み合わせが実現できる。

3.3 エンドユーザの利点

企業内システム部門が多言語翻訳プラットフォームを使い、多くの言語対のエンジンを活用した

新しいアプリを迅速に提供できるようになれば、エンドユーザも便利なツールが数多く使えるようになりメリットを享受できる。

また、インターネットポータルが希少な言語間の翻訳サービスを始めたり、これまでにないような言語処理の無料サービスを提供することも考えられ、エンドユーザによるインターネットの利用価値がさらに高まると考えられる。

4. 利用実験

4.1 実証実験システムからの検証

日中間の企業情報・知財情報調査を目的とする実用的 Web システムを構築し、次の機能から多言語翻訳プラットフォームの有用性を確認した。

- 中国語特許・法令のクロスリンガル検索
- 中国 - 日本間のメール本文の翻訳支援
- 中国定型文書の翻訳

特にクロスリンガル検索では単語分割、辞書検索、翻訳の機能を組合わせて効率的なインプリメントが可能であった。(図2)

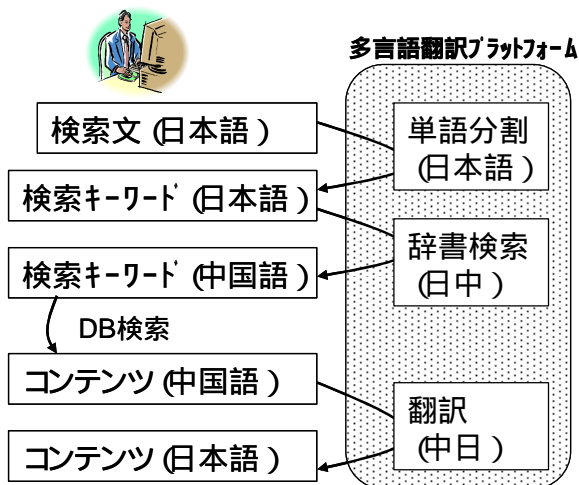


図2 クロスリンガル検索からの利用

4.2 各種言語からの検証

C++、VBA、Perl の各言語から翻訳プラットフォーム動作確認を行いいずれも正常に動作した。中国語 OS による中国から国際回線を経由したアクセステストを行い、実際の利用場面で実用的なレスポンスで利用できることがわかった。

また、翻訳プラットフォームへのアクセスが従来の DLL や Socket による翻訳機能呼出しに比べて、非常に簡単なコーディングで済むことを確認した。図3に Excel シートの一行に並んだ英語文を日本語に一括翻訳して隣のセルに結果を入れる

マクロを示す。翻訳を行うコードはオブジェクトの定義とメソッド呼出しの僅か2行(▶)である。

```
Private Sub CommandButton4_Click()
    Dim vRanges As Range
    Dim vRng As Range

    ▶ Dim trans As New clsWS_MultilingualServiceCl
    Set vRanges = Range(Range("b2"), Range("b11"))
    counter = 2
    For Each vRng In vRanges
        Cells(counter, 3) =
    ▶ trans.wsm_Trans_text(vRng, "ENJA", "", "", "", "", "")
        counter = counter + 1
    Next vRng
End Sub
```

図3 翻訳プラットフォームを呼出す Excel マクロ例

5. おわりに

本稿では、翻訳エンジンの出力(あるいは要素機能の出力)を標準化することにより自動翻訳の要素技術や訳例検索を Web サービスとしてユーザがシームレスに活用できる枠組みと、その設計・構築について述べた。そして多言語翻訳プラットフォームの枠組みがシステム開発の効率化、機械翻訳ビジネスの発展に寄与することに言及した。接続する翻訳エンジンは商用エンジンに限定されない。例えば世界中に無数に存在する実験用の自然言語ツールを翻訳プラットフォームに集約できれば、誰でも手軽に標準化された様々なツールを使うことが出来るようになり、自然言語研究の加速が期待できる。

しかしながら実際に多言語翻訳プラットフォームを運営する場合には、初期投資、運用保守、PR 活動などの負担元の調整など、技術面以外で解決すべき課題が多い。今後官民一体となった取り組みが必要であろう。

謝辞 本研究は平成 15~17 年度の総務省プロジェクト「国際情報通信ハブ形成のための高度 IT 共同実験」の一環として実施されたものである。

参考文献

- [1] Alan LaMont Pope. The CORBA reference guide: understanding the Common Object Request Broker Architecture. Addison-Wesley Longman Publishing Co., Inc., 1998.
- [2] SOAP 1.2
<http://www.w3.org/TR/soap12/>
- [3] DCOM
<http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnanchor/html/dcom.asp>