

# 多言語知識源の連携による留学生支援システムの提案

木野村 晋吾<sup>†</sup> 北川 大輔<sup>‡</sup> 桑原 和宏<sup>‡</sup>

<sup>†</sup>立命館大学大学院 理工学研究科

<sup>‡</sup>立命館大学 情報理工学部

## 1 はじめに

近年、海外から日本への留学生が増加しており、これらの留学生は必ずしも日本語を十分に使えるとは限らない。言語の壁や習慣の違いにより、異国で生活する留学生は多くの問題を抱えている。それに伴い、多言語間のコミュニケーションを支援する多言語アプリケーションの必要性が高まっており、同時に、多言語アプリケーションの開発を支援する取り組みも多く行われている。その一つに、多言語サービス基盤の言語グリッドプロジェクト [1] があり、機械翻訳や形態素解析、対訳などの様々な言語資源・言語サービスを提供し、開発者の支援を行なっている。その他にも、辞書や画像といった様々な種類の多言語知識源が、参照可能な形で一般に公開されている。多言語アプリケーションでは、これらの多言語知識源を利用し、参照可能な情報量を増加させ、ユーザの理解を補助することが重要である。しかし、収録されている知識や参照方法の異なる複数の多言語知識源に対して横断的に活用することが難しく、また、知識源が問い合わせとして対応している言語にも限りがあるため、対応していない言語によるユーザからのアクセスを考慮する必要もある。

そこで、複数の多言語知識源や機械翻訳サービスなどを連携させることにより、多くの情報に対して多言語アクセスを容易にする枠組みを提案する。この枠組みでは、複数の知識源への横断的検索や、検索語を知識源が対応する言語へ変換するといったように、多言語アプリケーションと多言語知識源の間の仲介を行なう。これにより、多言語アプリケーション開発者は、複数の知識源から言語を意識しないで情報を参照可能となる。また、提案する枠組みを利用し多言語知識源を連携させた例として、留学生がアパートで生活する上でよく出る質問とその回答 (FAQ) を集めたデータベースや住居に関する対訳辞書、住居設備の画像データベースなどを連携させて、生活上の情報を多言語で提供する生活情報提供システムについて述べる。

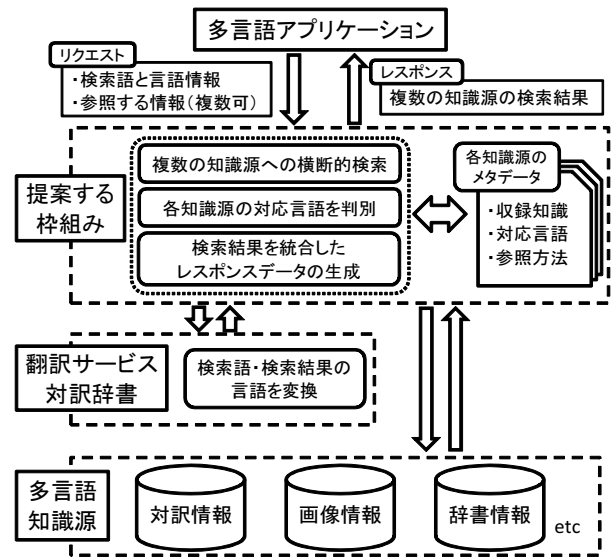


図 1: 枠組みを利用した多言語知識源の連携

## 2 多言語知識源の連携の枠組み

収録知識、対応言語、参照方法の異なる知識源を横断的に活用し、多言語アプリケーションと多言語知識源の間の仲介を行なう枠組みについて述べる。また、枠組みを利用した多言語知識源の連携を図 1 に示す。

### 2.1 複数の知識源への横断的検索

多言語アプリケーションが本提案の枠組みを介して知識源から情報の参照を行なうには、検索語とその言語情報、参照する情報といった内容のリクエストを枠組みに送信する。リクエストの内容に応じて複数の知識源に対して横断的に検索を行ない、それらの検索結果を統合し、レスポンスとして多言語アプリケーションへ返す。これにより、アプリケーション側から複数の知識源に直接アクセスする必要はなく、参照したい情報をリクエストとして枠組みに送るだけで、多くの情報を得ることが可能である。

### 2.2 リクエスト内容の対応言語への変換

知識源が問い合わせとして対応している言語には限りがあるため、多言語アプリケーション開発者が知識源に直接アクセスするような仕組みを作っても、知識源が対応していない言語でユーザから問い合わせをさ

“A Proposal of International Student Support System with Multilingual Knowledge Source Collaboration”  
Shingo KINOMURA<sup>†</sup>, Daisuke KITAGAWA<sup>‡</sup>, and Kazuhiro KUWABARA<sup>‡</sup>

<sup>†</sup>Graduate School of Science and Engineering, Ritsumeikan University

<sup>‡</sup>College of Information Science and Engineering, Ritsumeikan University

れると、情報を参照することができない。

そこで、この枠組みでは言語横断検索 (CLIR: Cross-Language Information Retrieval) [2] を行なう。検索語が検索対象の知識源に対応していない場合、翻訳サービスや対訳辞書などを介して検索語を対応する言語に変換したのち検索を行なう。また、検索結果として得た情報も必要に応じて言語の変換を行なう。これにより、本来知識源が対応していない言語による問い合わせであっても、知識源から情報を得ることが可能である。

### 3 枠組みを利用した留学生支援システム

提案する枠組みの機能を検証するために、生活上の情報を多言語で提供する留学生へ向けた生活情報提供システムを試作した。

#### 3.1 多言語知識源の構築

検証用に、賃貸管理会社から提供された、住居に関する FAQ と対訳の多言語データを基に、多言語知識源を構築した。これらのデータは、日本語・英語・中国語・韓国語により記述されている。賃貸 FAQ には“お風呂の排水が詰まった”などの、約 300 種の住居に関するトラブルとその解決方法が収録されており、賃貸対訳には、約 2500 種の住居に関する用語の対訳が収録されている。また、日本語でのアクセスにのみ対応している、住居設備の画像を収録したデータベースも同時に構築し、検索語を対応言語へ変換する機能を用いる。賃貸 FAQ、賃貸対訳、設備画像はそれぞれ個別の知識源として別サーバで構築しており、各知識源へのアクセス方法やレスポンスの内容はそれぞれ異なる。

#### 3.2 生活情報提供システム

本提案の枠組みを利用し、知識源から情報を参照する多言語アプリケーションの例として、3.1 節で述べた賃貸 FAQ、賃貸対訳、設備画像の三つの知識源に対して横断的活用を行ない、生活上の情報を多言語で提供する生活情報提供システムを構築した。このシステムでは、指定した言語の検索語に関する様々な生活上の情報を参照し、それらの結果をユーザに提供する。今回は、賃貸 FAQ と設備画像の情報を提供する。また、枠組みからのレスポンスには軽量のテキストベースのデータ交換フォーマットの JSON を用いる。これにより、JavaScript を始め多くのプログラミング言語でレスポンスの内容を取り扱うことが可能である。

生活情報提供システムにおける知識源の連携例を図 2 に示す。リクエストを枠組みに送信すると、検索語が検索対象の知識源に対応しているかどうかを判断される。対応している場合は、知識源に対して検索を行ない、対応していない場合は、2.2 節で述べたように、検索語を対応言語に変換したのち検索を行なう。図の例では、“air conditioner” という英語の検索語が送られてきた場合、設備画像の知識源は日本語にしか対応

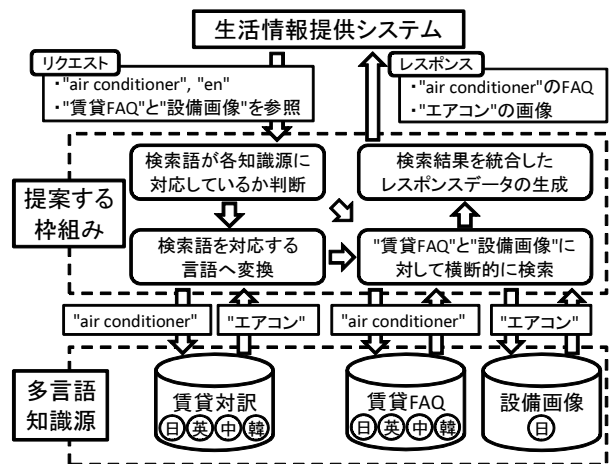


図 2: 生活情報提供システムにおける知識源の連携例

していないため、賃貸対訳を利用し“air conditioner”を“エアコン”という日本語に変換したのち検索を行っている。リクエストで指定した、賃貸 FAQ と設備画像の検索が完了すると、レスポンスとして生成した JSON 形式のデータを返す。これにより、生活情報提供システム側では、取得したレスポンスの内容について処理することで多言語情報を提供できる。

### 4 まとめ

本稿では、多言語知識源を連携させる枠組みと、それを用いて構築した留学生に生活上の情報を多言語で提供するアプリケーションについて述べた。今後の取り組みとして、インターネット上に公開されている知識源も含めて、知識源の連携が可能かを検証することがあげられる。また、各知識源から情報を参照する一連の手順は、現在のところ手動により行なっている。これを各知識源のメタデータから自動的に生成できるようにすることにも取り組んでいきたい。

### 謝辞

多言語データを提供して下さった株式会社ハウスセゾン中井俊彦氏に深謝いたします。

### 参考文献

- [1] Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration, *Proc. IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-100 (2006).
- [2] Fujii, A. and Ishikawa, T.: Japanese/English Cross-Language Information Retrieval: Exploration of Query Translation and Transliteration, *Computers and the Humanities*, Vol.35, No.4, pp.389-420 (2001).