

サービスコンピューティングに基づく多言語サービス基盤の実現 (092007004)

Multi-Language Service Infrastructure Based on Services Computing

研究代表者

石田 亨 京都大学大学院情報学研究科

Toru Ishida

Graduate School of Informatics, Kyoto University

研究分担者

松原 繁夫[†] 服部 宏充[†] 稲葉 利江子[†] 高崎 俊之[†] 林 冬恵[†] 中島 悠[†] 村上 陽平^{††} 田仲 正弘^{††}
アリフ ブラマントロ^{††} 林 良彦^{†††} 檜和 千春^{†††}

Shigeo Matsubara[†] Hiromitsu Hattori[†] Rieko Inaba[†] Toshiyuki Takasaki[†] Donghui Lin[†]
Yuu Nakajima[†] Yohei Murakami^{††} Masahiro Tanaka^{††} Arif Bramantoro^{††} Yoshihiko Hayashi^{†††}
Chiharu Narawa^{†††}

[†]京都大学大学院情報学研究科 ^{††}独立行政法人情報通信研究機構ユニバーサルコミュニケーション研究所
^{†††}大阪大学大学院言語文化研究科

[†]Graduate School of Informatics, Kyoto University ^{††}Universal Communication Research Institute,
National Institute of Information and Communications Technology

^{†††}Graduate School of Language and Culture, Osaka University

研究期間 平成 21 年度～平成 23 年度

概要

言語サービスオントロジーの標準化に関して、言語グリッドに登録されている具体的な言語資源・言語サービス群を対象として、オントロジー体系を詳細化し、これに立脚して言語サービスの API 体系の標準化の提案を行った。また、言語サービスの連携・評価機能の構築に関して、複合サービスの実行を監視し、複合サービスの QoS 制御を実行時に可能とする「サービススーパービジョン」機能を開発した。さらに、言語グリッドの国際展開に関して、タイ、インドネシアとの連邦制による言語グリッドの運営を実現し、Wikipedia 記事の多言語翻訳、NPO の多国籍スタッフのコミュニケーション支援、電子情報通信学会のアジア多言語配信をフィールドとした実証実験を行った。

Abstract

This project consists of three research issues: language service ontology standardization, language service composition, and international expansion of the language service infrastructure. First, we created ontology for language services and formed standard API for language services. Then, we implemented the service supervision framework for monitoring composite service execution and controlling QoS for composite services. Finally, we realized federated operation of the Language Grid with universities/research institutes in Thailand and Indonesia and conducted experiments including Wikipedia translation support, multilingual communication support for NPO international staff, and multilingual information sharing among Asian sections in IEICE.

1. まえがき

本研究は、「サービス指向の集合知 (Collective Intelligence)」の形成を、多言語サービスに特化して世界規模で実現することを目的とする。従来のコンテンツを集積した集合知である Wikipedia とは異なり、サービスを単位として集合知の形成を図り、多言語サービス基盤を、実際に実現しようとするものである。

平成 21 年度から平成 23 年度の 3 年間をかけて、「言語資源から言語サービスへ」という目標を実現するために、言語資源の API 体系を標準化する言語サービスオントロジーを多言語サービス基盤「言語グリッド」上で実現し、言語サービスの品質を保証する言語サービスの連携・評価機能を構築した。一方、言語グリッドを世界規模で展開するために、タイやインドネシアなどの大学や研究機関と連携して多言語サービス基盤の連邦制運営を実現した。さらに、言語グリッドを用いた多言語ツール群を開発し、アジアワイドな実証実験を行い、言語グリッド利用のベストプラクティスを創出した。

2. 研究内容及び成果

本研究の目標を達成するために、言語サービスオントロ

ジーの標準化と、言語サービスの連携・評価機能の構築、言語グリッドの国際展開という 3 つの課題に分けて研究を進めてきた。研究内容と成果の説明を図 1 に示す。

まず、言語サービスオントロジーの標準化の研究では、言語サービスの相互運用性を確保し、サービスの再利用性を高めるために、多言語サービス基盤「言語グリッド」に登録される言語サービスのインタフェースを体系化する必要がある。また、この体系化の基盤として、言語サービスの構成要素を規定する標準的な言語サービスオントロジーが求められる。この課題に対して、言語グリッドの 14 種のサービスインタフェースや公開されている言語解析ツールの API を分析することで、標準的と考えられる言語サービス・機能を抽出し、これらを利用するインタフェースを言語サービスオントロジーとして体系化した。これらにより、新たな言語資源のサービス化や言語サービスの連携を促進することが可能になる。さらに言語サービスの相互運用性を高めるため、言語資源に関する国際標準を適用した言語サービスの構成方法を実証的に示した。

次に、言語サービスの連携・評価機能の研究では、多言語翻訳の品質を保証するために、翻訳サービスや辞書・用例対訳サービスの連携を通じて複合サービスの実行管理

を行い、品質を制御する技術が必要であった。この課題に対して、まずは複合サービス実行時に言語サービスを選択する動的バインディング機能により 11 種の翻訳サービスと 54 種の辞書・用例対訳サービスの連携を実現した。さらに、言語サービスの選択だけでなく複合サービスのワークフローの実行制御を行うサービススーパービジョンの試作を行った。また、文脈に基づく訳語選択手法により 50% の翻訳文に対する品質を改善し、さらに品質評価サービスを用いた最良翻訳選択手法により翻訳品質を 5 段階評価で 4.0 まで向上させることを確認している。

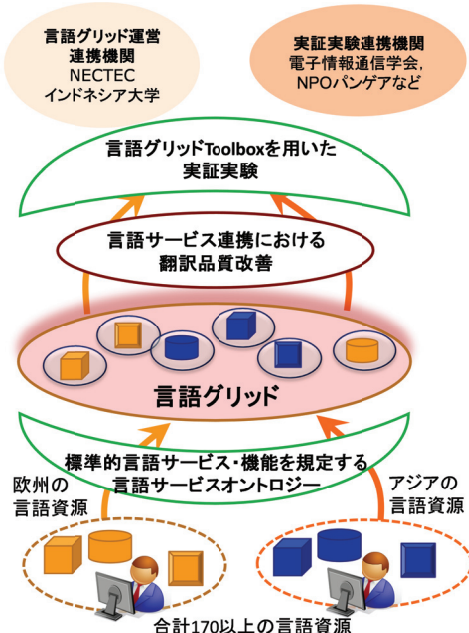


図1 研究成果説明図

最後に、言語グリッドの国際展開の研究では、アジアの大学や研究機関と共同して言語グリッドの連邦制運営が必要である。連邦制の運営を行うことにより、言語グリッドの参加グループの拡大と言語サービスの集積の加速を図る。さらに、言語グリッドに登録されている言語サービスに基づいて多言語ツール群を開発し、それを用いたアジアワイドな実証実験を実施し、言語グリッド利用のベストプラクティスを創出する。これを実現するため、言語グリッドの管理システムであるサービスマネージャを分散化し接続プロトコルを開発することで、京都、バンコク、ジャカルタによる連邦制運営を実現した。これにより言語グリッドの参加グループは 18 カ国合計 140 を超え、中国科学院や CNR、DFKI、NII といった研究機関や、シュツトガルト大学、プリンストン大学、清華大学、そして多くの日本の大学、NPO/NGO や公的機関などが参加することになった。NTT や東芝、沖電気、Google といった企業も参加しており、無償で機械翻訳サービスなどを提供している。また、共有されている言語サービスは機械翻訳、辞書、用例対訳、形態素解析、音声合成など 20 種合計 170 を超えている。さらに、これらの言語サービスを簡単に利用できる言語グリッド Toolbox という多言語ツール群を開発し、Wikipedia 記事の多言語翻訳、NPO パンゲアによる国際子供活動を運営する多国籍スタッフのコミュニケーション支援、電子情報通信学会のアジア多言語配信をフィールドとした実証実験を実施している。

3. むすび

本研究は、サービスコンピューティング技術に基づいて、標準化されたオントロジーを用いた言語サービスの収集、

共有を支援するサービス指向の多言語基盤「言語グリッド」を実現し、多言語サービスの品質を保証するためのサービス連携・評価技術を開発した。また、言語グリッドの連邦制の運営と多言語ツール群の実現によって、3 種類のアジアワイドな実証実験を行った。これらの研究成果に基づいて、言語サービスタイプの追加と言語サービスを用いた多言語コミュニケーション支援の研究はますます活発化している。また、本研究プロジェクトの終了後も、言語グリッドの研究開発と運営を継続しており、言語サービスの増加と参加グループの拡大に努めている。

また、これまでの活動により、我々が提唱した「言語資源から言語サービスへ」という言語グリッドの方向性が、欧州、米国の言語資源研究者の間で共有され始めている。米国では、自然言語処理、情報検索、機械翻訳、音声、セマンティックウェブなどの分野で、これまで個別に作成されてきた言語資源を、分野を超えて再利用するプロジェクト SILT (Sustainability Interoperability for Language Technology) が進められてきた。SILT の次期プロジェクトである National Science Foundation (NSF) 「The Language Application Grid: A Framework for Rapid Adaptation and Reuse」(2012~2015) は、言語グリッド基盤ソフトウェアの利用と言語サービスの標準化に関する共同研究について、言語グリッド側と合意した。言語サービスを世界規模で共有するために、欧米やアジアとの協力を今後も継続して進めていく予定である。

【誌上発表リスト】

- [1] Masahiro Tanaka, Yohei Murakami, and Toru Ishida, "Service Supervision for Runtime Service Management". Toru Ishida (ed.) The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability, pp. 69-83, Springer. (2011 年 8 月)
- [2] 石田 亨、村上陽平、稲葉利江子、林冬恵、田仲正弘、"言語グリッド：サービス指向の多言語基盤"、電子情報通信学会論文誌 Vol. J95-D No.1, pp. 2-10 (2012 年 1 月)
- [3] Yoshihiko Hayashi, Bora Savas, Monica Monachini, Claudia Soria, Nicoletta Calzolari, "LMF-aware Web Services for Accessing Semantic Lexicons", Language Resources of Evaluation (2012 年 3 月 30 日)

【申請特許リスト】

- [1] 高崎俊之、石松昌展、石田 亨、情報処理システム、サーバ、端末装置、情報処理方法、及びプログラム、日本、2010 年 2 月 4 日
- [2] 林 冬恵、石田 亨、村上陽平、田仲正弘、サーバ装置、情報処理方法、およびプログラム、日本、2010 年 5 月 6 日
- [3] 田仲正弘、林 冬恵、村上陽平、石田 亨、情報システム、コアサーバ、サービス提供方法、およびプログラム、日本、2010 年 7 月 2 日

【報道発表リスト】

- [1] "情通機構など、多言語支援機能を利用しやすくしたツールボックス開発"、朝日新聞(Web ニュース)、2011 年 1 月 15 日
- [2] "多言語交流 ソフトで支援"、京都新聞 (朝刊・25)、2011 年 1 月 21 日
- [3] "「言語グリッド」の世界展開"、産経新聞 22 面、2011 年 3 月 11 日

【本研究開発課題を掲載したホームページ】

言語グリッドポータルサイト <http://langrid.org>