

6-4 言語グリッド: サービス指向集合知による多言語基盤

6-4 *The Language Grid: Multi-Language Infrastructure Based on Service-Oriented Collective Intelligence*

村上陽平

MURAKAMI Yohei

要旨

言語グリッドは、異文化コラボレーションを支援するための、サービス指向の多言語基盤である。辞書などのデータや機械翻訳などのソフトウェアが言語資源としてCDやダウンロードサービスで配布されている現状から、インターネットに接続すればすぐに利用できる言語サービスへと転換し、利用者が異文化コラボレーションの環境に合わせて言語サービスを容易に組み合わせられるようにすることが目的である。この論文では、我々が開発した、言語サービスの収集と共有、連携を可能とする「基盤ソフトウェア」と、基盤ソフトウェアに登録された言語サービスを容易に利用可能とする「異文化コラボレーション環境」について説明し、フィールドでの利用事例を報告する。

The Language Grid is a service-oriented multi-language infrastructure for supporting intercultural collaboration. Language resources like dictionary data and machine translation software that are traditionally distributed with CD copies and are downloaded should be transformed to language services that can be easily available through the Internet and allow users to combine them according to intercultural collaboration fields. In this paper, we introduce the infrastructure software that enables collecting, sharing, and combining language services, the intercultural collaboration environment that enables easy usage of language services registered in the infrastructure software, and several typical use cases.

[キーワード]

サービス指向, 多言語基盤, サービスグリッド, 言語グリッド, 異文化コラボレーション

Service-oriented, Multi-language infrastructure, Service grid, Language grid, Intercultural collaboration

1 まえがき

インターネットは世界の人々を繋いだと言われるが、言語の壁は依然として存在している。インターネット上の言語人口は多様化し、標準言語がない状態ともいえる。Global Reachの調査によれば、インターネット上の英語人口は35.2%、アジア言語人口が26%、欧州言語人口が28%である。Webで様々な情報が共有されているが、それらの情報を知るためには、多くの言語の理解が必要となり非常に困難である。また、インターネット上には多数の言語資源(データ及びソフト

ウェア)が存在しているが、専門家でなければ異文化コラボレーション活動の現場で利用することは難しい。複雑な契約や知的財産、データ構造やインタフェースの多様性が言語資源の利用を困難にしている。

このような問題には、各ユーザが個別に振る舞うよりも集合として知的に振る舞う集合知^{[1][2]}が有力な解決策である。近年、Wikipediaのように、世界規模でWeb上に集合知を形成する取り組みが人類の知識基盤を生み出している。現在進んでいる集合知の形成が、文章、写真、動画などを組み上げていくコンテンツ指向の集合知である

のに対し、本研究では、機械翻訳や音声認識などのサービスを組み上げるサービス指向の集合知を提案する。特に、異文化コラボレーションを支援するための言語サービスの集合知基盤である「言語グリッド」について説明する[3]。言語グリッドでは、エンドユーザが研究者や専門家の提供する既存の言語サービスを組み合わせ、さらに自分自身の言語サービスを追加することで、自分の目的に合わせて新しい言語サービスを作成することが可能である。本論文では言語グリッドの構築時に取り組んだ下記の3つの課題について述べる。

- ・サービス指向の多言語基盤の構築：言語サービスを蓄積し共有するには、標準インタフェースを持つ原子サービスに基づいてサービスを連携する基盤ソフトウェアが必要である。さらに、利用者がそれらの言語サービスを用いて異文化コラボレーションを支援するアプリケーションシステムを簡単に開発できなければならない。
- ・運用モデルの制度設計：サービス指向集合知には、多様なステークホルダーが存在する。利用者はそれぞれ異なる要件を持ち、提供者も異なるポリシーを持つ。運営者はこれらの異なるステークホルダーを協調させるよう、両者のインセンティブを考慮した運用モデルを設計しなければならない。
- ・ユーザ参加型デザインの実践：提供される言語サービスが多ければ多いほど、利用者はそのサービスによる利益を享受できる。つまり、サービス指向の集合知を形成するには、利用者とコミュニティを積極的に参加させることが必要である*1。

言語グリッドの研究は言語サービスを対象としたものであるが、サービスコンピューティングの一般的な課題を数多く含んでいる。例えば、オープンな環境においてサービスの蓄積と共有をいかに行うか、利用者やそのコミュニティによる新たなサービスの作成をいかに支援するか[5]等が問題としてあげられる。

本論文では、まず言語グリッドの基盤ソフトウェアについて述べ、その後、そのソフトウェアを用いた言語グリッドの運用モデルについて説明する。最後にユーザ参加型デザインについて、ローカルコミュニティおよびグローバルコミュニ

ティでの言語グリッドの実践事例を紹介する。

2 言語グリッドアーキテクチャ

2.1 設計思想

言語グリッドは、集合知のアプローチを取っている。即ち、専門家や様々な利用現場のユーザが開発した言語資源を共有し利用できる環境として設計されている(図1)。言語グリッドの特徴は、言語資源をサービスの形で共有することである。そこには、言語グリッド運営者、サービス提供者、サービス利用者の3種のステークホルダーが存在する。言語グリッド運営者は、言語グリッドを管理し、言語サービスの実行を制御する。サービス提供者は、機械翻訳や形態素解析、辞書などの言語資源を言語サービスとしてラッピングし言語グリッドに登録する。サービス利用者は登録された言語サービスを呼び出して異文化コラボレーション活動に利用する。

このように言語グリッドは異なる組織から提供される言語サービスを結合するプラットフォームである。これまでも言語処理プログラムを結合しようとする試みとしてDFKIのHeart of Gold[6]やIBMのUIMA[7]が存在したが、主に研究開発者のためのプラットフォームで、共有データに対して、所有する多様な言語処理プログラムをパイプライン的に適用することができる。一方、言語グリッドはアプリケーションが言語資源を利用するためのプラットフォームで、サービス指向アーキテクチャに基づいて知財を管理することに焦点を当てている。このように目的が直交するため、DFKIのHeart of Goalと言語グリッドをシステム的に連結する共同研究も行われた[8]。

2.2 システムアーキテクチャ

図2に示すように、言語グリッドは概念的に以下の4層から構成される[9]。最下層のP2Pサービスグリッドは、コアノードとサービスノードという2種類のノードを接続することを目的としている。コアノードはサービスの登録情報を管理し、サービスのアクセス制御を行い、サービ

*1 集合知の成長は利用者の自発的な努力によるものとされている[4]。

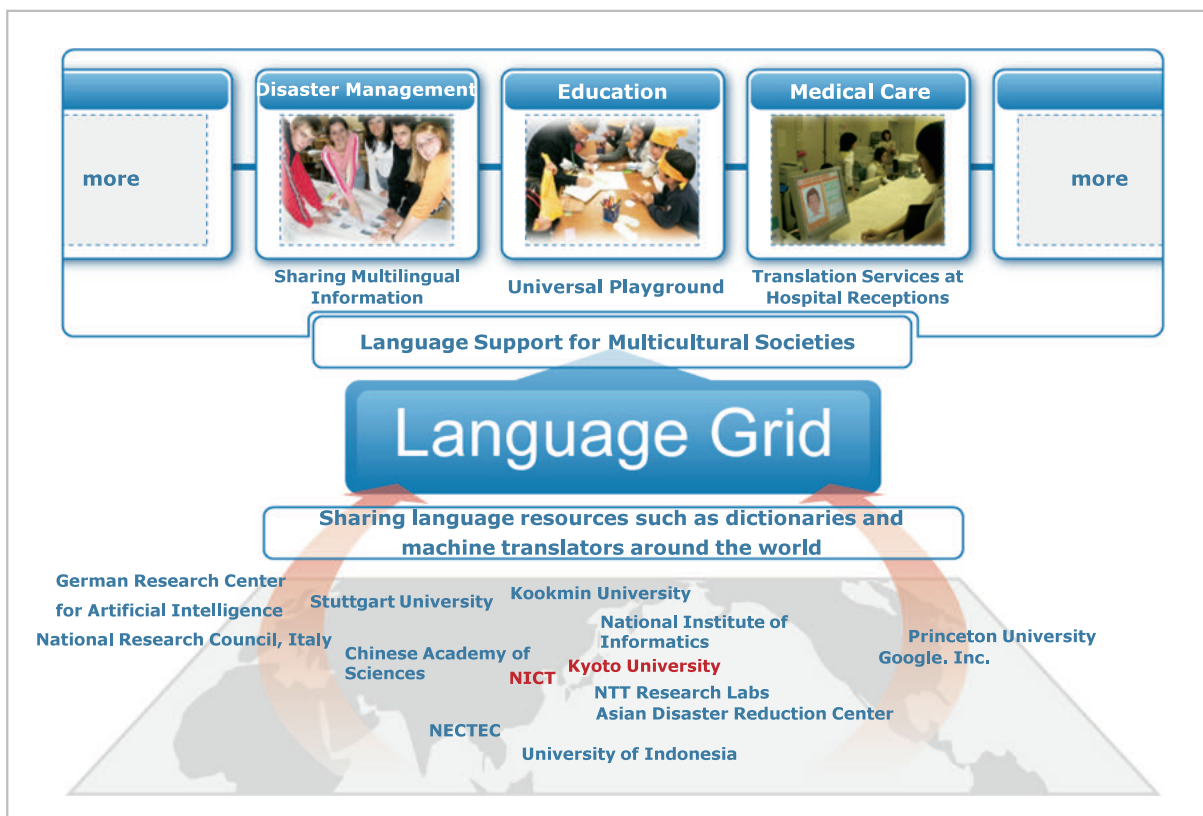


図 1 言語グリッド

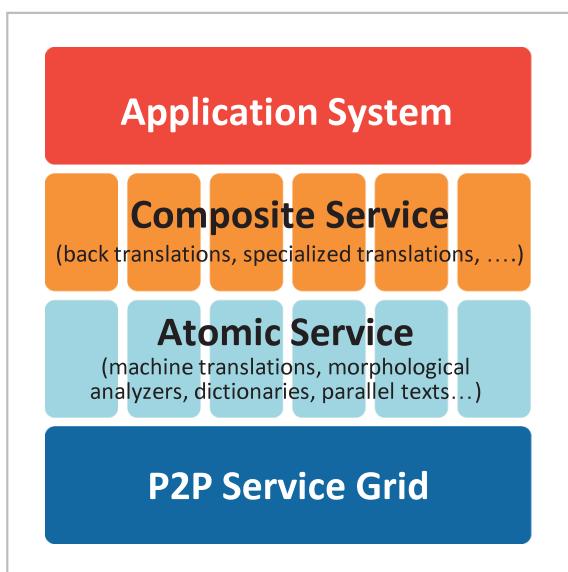


図 2 言語グリッドの階層

スを連携させる。一方、サービスノードには、サービス実体となる言語資源とそのラッパーが配備される。

原子サービスは、個々の言語資源に対応した Web サービスであり、サービスノード上で提供

される。例えば、機械翻訳や形態素解析、辞書、用例対訳が典型的な言語資源である。これらの言語資源は標準化されたサービスインタフェースに基づいてラッピングされる。既に、様々な言語データや言語処理プログラムのサービスインタフェースを階層的に標準化するためのオントロジー体系が提案されている [10]。

複合サービスは、ワークフローによって原子サービスを合成したものである [11]。ワークフローは WS-BPEL によって記述され、BPEL 実行エンジンによって解釈、実行される [12]。言語ドメインでは、翻訳結果を再度入力言語に翻訳し直す折り返し翻訳サービスや、専門用語辞書サービスと翻訳サービスを連携させた専門翻訳サービスといった多様な複合サービスが構築される。

応用システムは、異文化コラボレーションを支援するために、各現場に特化して構築されたアプリケーションである。複合サービスや原子サービスを利用することで、応用システムはユーザとのインタラクションを行うユーザインタフェースの開発に集中でき、開発コストの削減が可能になる。

この言語グリッドの4層構成を支えるP2Pサービスグリッドのシステム構成を図3に示す[13]。サービス提供者は、言語資源をラッピングした後、Webサービスのインタフェース記述であるWSDLファイルとサービスの著作権情報、ライセンス情報、アクセス制約をサービスマネージャ (Service Manager) に登録する。サービスマネージャは、WSDLファイルを取得すると、インタフェース情報とエンドポイントのURLを抽出し、同じインタフェースの仮想エンドポイントをサービススーパーバイザ (Service Supervisor) 上に生成する。仮想エンドポイントの目的は、サービスへの直接のアクセスを禁止し、サービスの利用状況をモニタリングするとともに、指定されたアクセス制約に基づいて、サービスへのアクセスを制御することである。

サービスを利用するときには、応用システムから仮想エンドポイントにSOAPリクエストを送りサービスを呼び出す。サービススーパーバイザは、そのリクエストを受け取ると、サービス登録時に設定されたアクセス制約を満たしているかどうか検証する。満たしていれば、サービススーパーバイザはそのサービスの実際のエンドポイントから取得しサービスにアクセスする。サービスが複合サービ

スの場合は、リクエストがサービスワークフロー実行系に送られ、対応するワークフローが実行される。サービスワークフロー実行系は、ワークフローに基づいて新しいリクエストをサービススーパーバイザに送信する。なお、サービスからのレスポンスはアクセスログに蓄積され、アクセス制約の充足性の検証や、サービス利用のモニタリングに利用される。

アクセスログに蓄積されるこれらの情報や、サービス提供者から登録されたサービスの情報などは、グリッドコンポーザを介して他のコアノードにも共有される。これにより、サービス利用者は、複数あるコアノードのいずれからも同等のサービスを受けることが可能である。

3 言語グリッドの運営

3.1 中央集権型運営モデル

言語グリッドの運営を開始するにあたり、ステークホルダーであるサービス提供者、サービス利用者の要件を収集し、非営利運営に限定した中央集権型運営モデルを構築した。中央集権型運営モデルでは、単一組織で言語グリッドを運営し、全ての利用組織はこの運営組織と覚書を締結する。サービス提供者は提供サービスへのアクセス

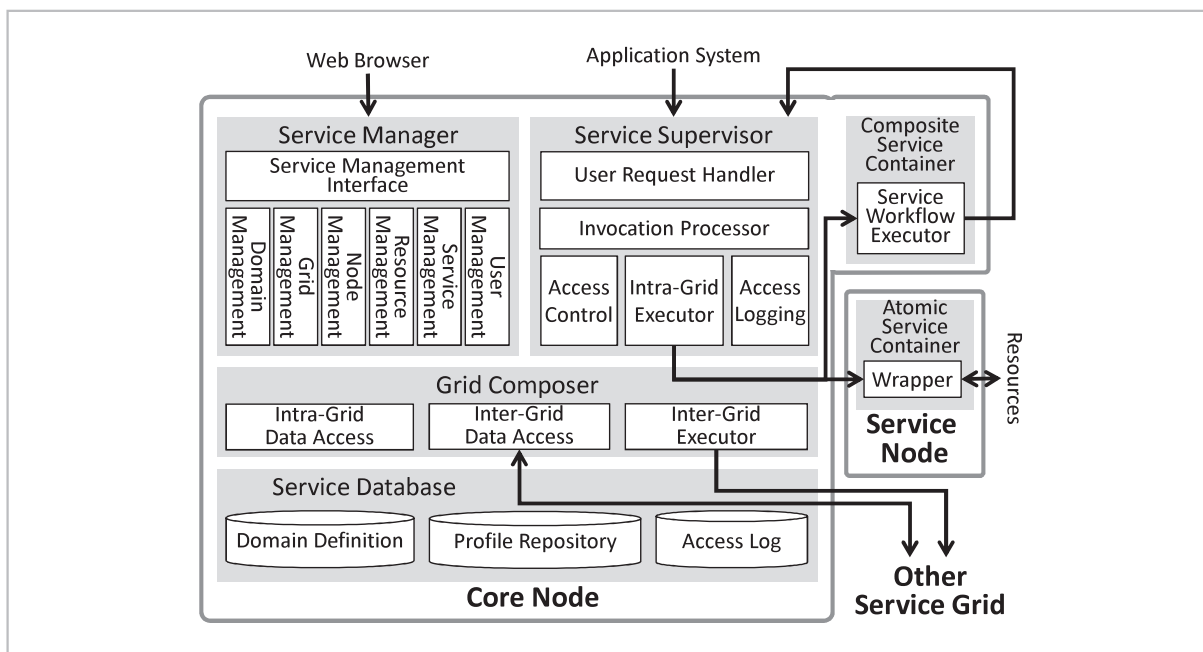


図3 P2P サービスグリッドのシステム構成

を完全に制御可能である。たとえば、サービス利用者の利用統計情報（アクセス回数、アクセスログ、データ転送量等）をモニタリングでき、利用可能なユーザの選択や利用制約の設定も可能である。さらに、不正アクセスを防ぐために、サービス利用者はツールを公開して不特定多数のエンドユーザからのアクセスを可能にしてはならない。つまり、ツールを公開する場合は、各エンドユーザをサービス利用者が同定できる必要がある [14]。この結果、現場で異文化コラボレーションに直面している NPO や NGO、自治体での利用は予想に反して少なく、大学での研究利用が多くなっている。図 4 に言語グリッドの利用組織の推移を示す。現在は 16 カ国 139 組織が覚書に署名している。参加組織は、例えば、中国科学院や CNR、DFKI、NII といった研究機関や、シュツットガルト大学、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 や公的機関などである。NTT や東芝、沖電気、Google といった企業も参加し無償で機械翻訳

サービスなどを提供している。

3.2 連邦制運営モデル

既存の中央集権的な非営利運営モデルでは、2つの課題がある。1つ目は、言語グリッドの利用組織のうち、国内の利用組織が70%以上を占めていることである。2つ目は、さらにその利用組織の大半が大学であることである。この課題に対処するために、複数の運営組織による連邦制での運営モデルの設計を行った (図 5)。連邦制運営モデルでは、単一の運営組織だけではなく、複数の運営組織が連携して運営を行うために、連携運営組織と連携利用組織というロールを導入している。連携運営組織とは、同一の覚書を用いて自らの言語グリッドを別途運営している利用組織のことをいう。また、連携利用組織とは、連携運営組織と覚書を締結した利用組織のことを言い、運営組織とは直接覚書を締結していない利用組織である。この連携利用組織によるサービスの利用を許可することで、単一の運営組織では接触できなかった利用組織にサービスの提供が可能になり、またその利用組織によって提供される言語サービスを収集することが可能になる。

もう一方の課題に対しては、言語グリッドの利用目的や利用方法の範囲を拡張し、選択式にしている。提供者はサービス登録時に、許可する利用目的と利用形態を設定し、利用者は利用時に利用目的や利用形態を送信する。システム上でお互いのマッチングを取りアクセス制御を行っている。これにより、営利企業の研究利用やサーバ型でのツール公開が可能になり、利用方法が多様化する

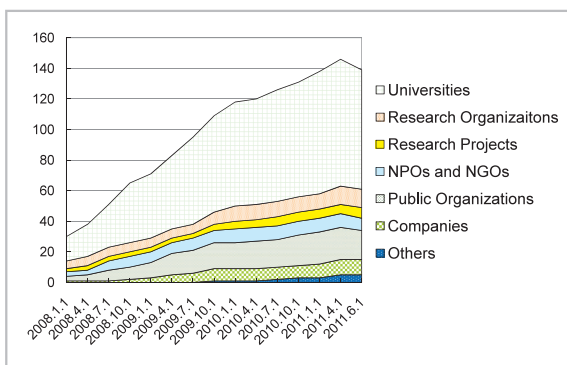


図 4 言語グリッド参加者数

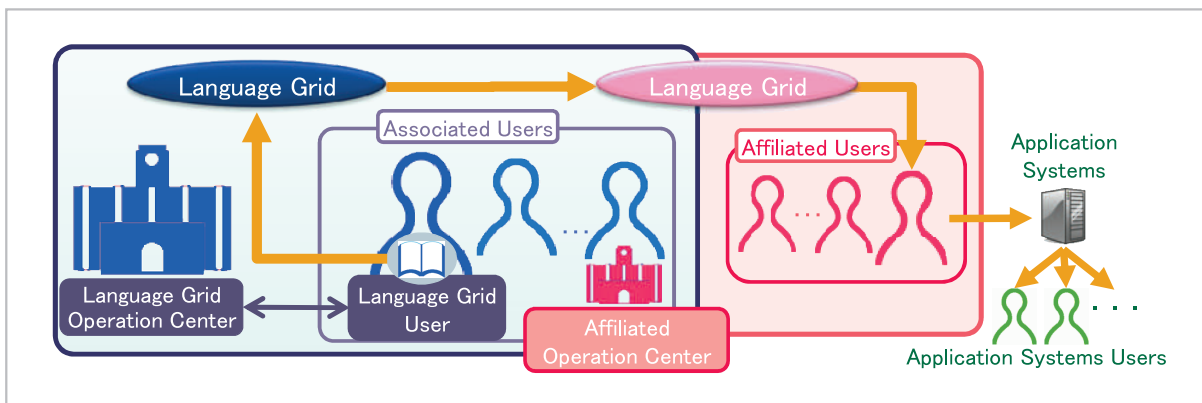


図 5 連邦制運営モデル

ことで利用組織の種類が増加が期待される。

2011年2月には、タイのNECTECが言語グリッドオペレーションセンターをバンコクに立ちあげ、京都大学のオペレーションセンターと連邦制運営を開始した[15]。その結果、言語グリッド(京都、バンコク)に登録された言語サービスは現在120を超え、多様な原子・複合サービスが、Translation、Bilingual Dictionary、Parallel Text、Morphological Analysis、Text-to-Speechなど20種のサービスタイプに分類され共有されている。

4 言語グリッドの利用

4.1 異文化コラボレーション支援環境

集合知のアプローチでは、利用者の自発的な活動を通してのみプラットフォームが成長していく。サービス提供者がサービスを提供すればするほど、サービス利用者はますますサービスの恩恵

を受けることができる。したがって、我々は言語グリッドの開始当初から潜在的なサービス利用者となりうるNPOなどと協力し、参加型設計のアプローチを取ってきた。これらのNPO/NGOや学校、他の非営利組織などは、翻訳品質を改善するために、自分用の辞書や用例対訳を作成しサービスとして提供し、言語グリッドを通してそれを通常の翻訳サービスに組み合わせて利用している。本研究では、特に防災や教育、医療の分野に焦点を当て、各分野にカスタマイズされた言語サービスを用いて異文化コラボレーション支援ツールの開発を行っている。本節では、分野ごとの異文化コラボレーション支援ツールの開発コストを削減するために開発した、汎用かつカスタマイズ可能な異文化コラボレーション支援ツールについて2つ紹介する。

まず言語グリッド Playground は京都大学の学生チームによって開発された応用システムで、言語グリッド上の様々な言語サービスに Web ブラ



図6 言語グリッド Playground

ウザを通じてアクセスすることができる。Playgroundには、原子サービスを利用するためのBasicコンポーネント、原子サービスを組み合わせた複合サービスを利用するためのAdvancedコンポーネント、異文化コラボレーション活動への応用に特化したCustomizedコンポーネントがある。Customizedコンポーネントは、応用に合わせてBasicコンポーネント、Advancedコンポーネントを組み合わせて構築される。図6に川崎市立富士見中学校用に開発された多言語チャットツールのCustomizedコンポーネントを示す。

一方、言語グリッドToolboxは、コミュニティにおける異文化コラボレーションを支援するモジュール群であり、多言語BBSなどのコミュニケーション支援機能を持つ。また、XOOPSと呼ばれるオープンソースソフトウェアのCMSに基づいて提供されており、各コミュニティが必要

に応じて拡張できる。現在、NPOパンゲアは、言語グリッドToolboxを利用して多言語コミュニティサイトを構築している(図7)。

4.2 ローカルコミュニティでの利用

在日外国人の増加に伴い、医療の現場においても、十分に日本語を話すことができない外国人患者との対話が大きな問題となっている。医療現場の場合、病状、薬、保険制度などが、医療従事者と患者の双方で正しく伝わらなければならない。京都では医療通訳ボランティアが同行する支援が行われており、その需要は増大している。

そこで、用例対訳を利用し、医療従事者と患者間の対面でのコミュニケーションを支援する多言語医療受付支援システムM³(図8)が、和歌山大学と多文化共生センターきょうとにより開発された[16]。医療現場、特に医療受付時に高頻度で利用される用例が必要となるため、医療用例収集



図7 言語グリッドToolboxの活用事例

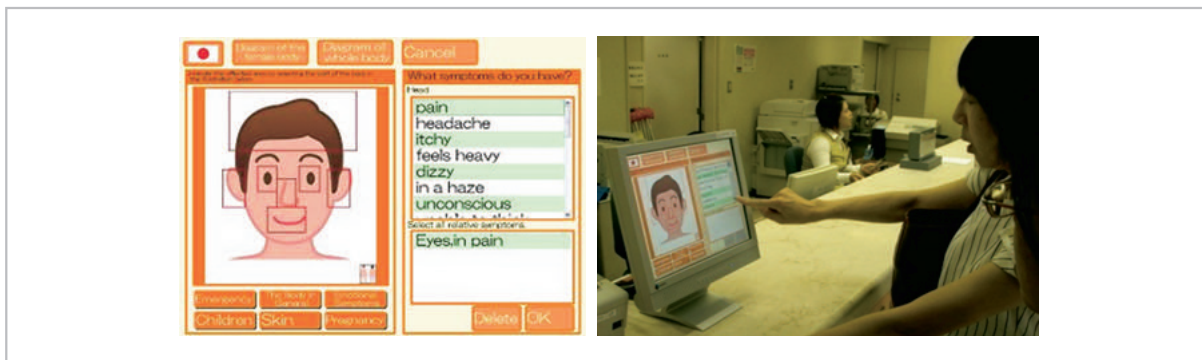


図8 多言語医療受付支援システムM³

システム TackPad が開発され、医療通訳ボランティアによる用例対訳の収集が行われている。

現在、M³ は、京都市立病院、京都大学医学部附属病院、洛和会音羽病院、東京大学医学部附属病院に導入され、多言語受付の支援が行われている。また、病院に行く前の医療支援を目的とした Web 版 M³ やモバイル版 M³ の公開も行われている。

4.3 グローバルコミュニティでの利用

Wikipedia は、誰でも記事を作成・編集できるため、約 270 もの言語により情報が共有されている。これらの記事はそれぞれの文化を背景に執筆されているため、異文化の相互理解のための知識の宝庫と言える。しかしながら、その内訳を調べると、英語では 354 万本の記事があるのに対し、日本語では 73 万本、タイ語では 6 万本など言語によって記事の数に大きな偏りがある。知識の翻訳を加速するためには、翻訳に関する議論が可能な多言語掲示板が必要である。

そこで Wikimedia 財団と共同で、言語グリッドを応用した多言語掲示板を MediaWiki 上に開発した^{*2}。この多言語掲示板を用いれば、世界中の Wikipedia ボランティアは、記事の翻訳のために多言語での質問応答を行うことができる。

実現方法としては、まず、MediaWiki 上に、言語グリッドへのアクセス手段を提供する言語グ

リッドエクステンションを開発した (図 9)。次に、これを利用し、Wikimedia 財団が開発した単言語の掲示板『Liquid Thread』を拡張した多言語掲示板『Multilingual Liquid Thread』を開発した。Multilingual Liquid Thread は、記事ごとに多言語用語集を作成できるため、記事ごとに機械翻訳をカスタマイズし、翻訳精度を向上させることができる。

5 むすび

本研究では、言語資源の知的財産権を保護しつつ、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を実現した。本研究の主な貢献は下記の 3 点である。

- ・ **サービス指向の多言語基盤の構築:** 言語グリッドは、P2P サービスグリッド、原子サービス、複合サービス、アプリケーションシステムといった 4 つの階層からなる。P2P サービスグリッドにより、コアノード間で情報共有が行われることで、サービス利用者はいずれのコアノードからも同じサービスを利用でき、サービス提供者はどのコアノード上でも統一したアクセス制御を行うことができる。
- ・ **運営モデルの制度設計:** 言語グリッドにおいて、異なるステークホルダーを協調させるための運営モデルを提案した。この運営モデルは、サービス利用者のインセンティブをサービス提供者のインセンティブに合致するように設計されている。さらに言語サービスのアクセシビリティを向上させるために、複数の運営組織による連携を可能にする連邦制運営モデルを提案した。
- ・ **ユーザ参加型デザインの実践:** サービス指向のアプローチと汎用の異文化コラボレーション支援ツールがユーザ参加型デザインを加速することを示した。実例として、利用者である学校や NPO が、言語グリッドプレイグラウンド、言語グリッド Toolbox を用いて専用の異文化コラボレーション環境を開発した事例を述べた。

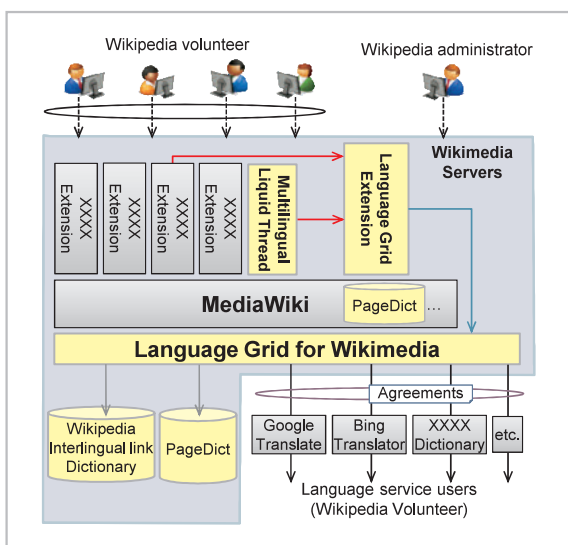


図 9 Multilingual Liquid Thread のシステム構成

*2 MediaWiki は Wikipedia など、Wikimedia 財団が提供するサービスのプラットフォームである。

言語グリッドは、利用者の目的に合わせた多言語環境を構築するためのサービス指向の多言語基盤である。各大学や研究機関、企業等が提供している言語サービスを利用者が自由に組み合わせることを可能にする。その結果、各地域の学校の多言語支援、商店街のコミュニティの支援等の活動に利用されている。例えば、世界中の子ども達が描いた災害安全マップをインターネット上で共有し、防災協働学習を支援するシステム CoSMOS (Collaborative Safety Maps on Open System) などが開発されている [17]。

言語グリッドは、異文化コラボレーションを支援するための、サービス指向集合知に基づく多言語基盤であるが、その基盤ソフトウェアや運用モデルは言語に特化したものではない。新たにサービスのインタフェースを定義することで、他のドメインへの応用が可能である。今後は、大規模な

サイエンスデータやアーカイブデータを対象としたデータサービスや、大規模データを分析するためのサービスを体系化することで、サービス指向集合知のアプローチに基づき、ビッグデータの利活用を促進するサービス基盤の開発に取り組んでいく予定である。

謝辞

本研究について、研究動機を与えて頂き、プロジェクトを率いて頂いた京都大学石田教授、日々の言語グリッドの運営を担っている言語グリッド京都オペレーションセンター (京都大学社会情報学専攻)、バンコクオペレーションセンター (NECTEC)、そして言語グリッド利用組織を支援している言語グリッドアソシエーションに感謝する。

参考文献

- 1 T. Gruber, "Collective Knowledge Systems: Where the Social Web meets the Semantic Web," *Journal of Web Semantics*, Vol. 6, No. 1, 2008.
- 2 P. Levy, "Collective Intelligence: Mankind's Emerging World in Cyberspace," Cambridge, MA: Perseus Books, 1999.
- 3 T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96–100, keynote address, 2006.
- 4 A. Weiss, "The Power of Collective Intelligence," *networker*, Vol. 9, No. 3, pp. 16–23, 2005.
- 5 M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-Oriented Computing: a Research Roadmap," *Int. J. Cooperative Inf. Syst.* Vol. 17, No. 2, pp. 223–255, 2008.
- 6 U. Callmeier, A. Eisele, U. Schafer, and M. Siegel, "The Deep Thought Core Architecture Framework," *LREC 2004*, pp. 1205–1208, 2004.
- 7 D. Ferrucci and A. Lally, "UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment," *Natural Language Engineering*, Vol. 10, pp. 327–348, 2004.
- 8 A. Bramantoro, T. Tanaka, Y. Murakami, U. Schäfer, and T. Ishida, "A Hybrid Integrated Architecture for Language Service Composition," *IEEE International Conference on Web Services (ICWS-08)*, pp. 345–352, 2008.
- 9 Y. Murakami, and T. Ishida, "A Layered Language Service Architecture for Intercultural Collaboration," *International Conference on Creating, Connecting and Collaborating through Computing (C5-08)*, 2008.
- 10 Y. Hayashi, T. Declerck, P. Buitelaar, and M. Monachini, "Ontologies for a Global Language Infrastructure," *Proc. of ICGL2008*, pp. 105–112, 2008.
- 11 R. Khalaf, N. Mukhi, and S. Weerawarana, "Service-Oriented Composition in BPEL4WS," *Proceedings of the 2003 World Wide Web Conference*, 2003.
- 12 T. Andrews, F. Curbera, H. Dolakia, J. Goland, J. Klein, F. Leymann, K. Liu, D. Roller, D. Smith, S. Thatte, I. Trickovic, and S. Weerawarana, "Business Process Execution Language for Web Services," 2003.

- 13 Y. Murakami, M. Tanaka, D. Lin, and T. Ishida, "Service Grid Federation Architecture for Heterogeneous Domains," International Conference on Services Computing (SCC-12), 2012.
- 14 T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa, "A Non-Profit Operation Model for the Language Grid," International Conference on Global Interoperability for Language Resources, pp. 114–121, 2008.
- 15 石田 亨, 村上陽平, "サービス指向集合知のための制度設計," 電子情報通信学会論文誌 D, Vol. J93-D, No. 6, pp. 675–682. 招待論文, 2010.
- 16 宮部真衣, 吉野 孝, 重野亜久里, "外国人患者のための用例対訳を用いた多言語医療受付支援システムの構築," 電子情報通信学会論文誌 D, Vol. J92-D, No. 6, pp. 708–718, 2009.
- 17 Y. Ikeda, Y. Yoshioka, and Y. Kitamura, "Intercultural Collaboration Support System Using Disaster Safety Map and Machine Translation," Culture and Computing, Lecture Notes in Computer Science 6259, Springer, 100–112, 2010.

(平成 24 年 6 月 14 日 採録)



むらかみ ようへい
村上陽平

ユニバーサルコミュニケーション研究所
情報利活用基盤研究室主任研究員
博士 (情報学)
サービスコンピューティング、マルチ
エージェントシステム
yohei@nict.go.jp