

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：34315

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500193

研究課題名(和文)異文化コラボレーションのためのセマンティックギャップ解消の研究

研究課題名(英文) Bridging semantic gap for multi-cultural collaboration

研究代表者

服部 文夫 (Hattori, Fumio)

立命館大学・情報理工学部・教授

研究者番号：80388131

交付決定額(研究期間全体)：(直接経費) 4,000,000円、(間接経費) 1,200,000円

研究成果の概要(和文)：異文化コラボレーションにおけるセマンティックギャップ(ユーザの背景知識や状況の違いによって生じるコンテンツに対する理解の差異)を解消するために、対象とするコンテンツのセマンティクスを記述したメタデータを翻訳するシステムを開発した。

Web文書からのメタデータの自動生成方法、オントロジー・マッピングによるメタデータ翻訳方法、および異言語のオントロジー間のマッピング方法を提案した。

料理レシピを例にとり、日英言語間のプロトタイプを構築して、提案手法の実現可能性を明らかにした。

研究成果の概要(英文)：To bridge the semantic gap in multi-cultural collaboration, a meta-data translation system is developed. A method of automatic meta-data generation from Web documents, meta-data translation mechanism using ontology mapping, and mapping algorithm between multi-language ontologies are proposed. Usefulness of the proposed method is demonstrated by building a prototype system between Japanese and English using cooking recipe contents as an example.

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：セマンティックWeb オントロジー 異文化コミュニケーション 知識ベース

1. 研究開始当初の背景

異文化のユーザ間のコミュニケーションの支援を狙いとして、言語グリッドを用いたチャットシステムや電子会議システムの研究がすすめられている。しかし、これらはいずれもチャット内容や文書そのものを翻訳して提供することを実現しているにすぎない。また、多言語での共同作業の支援を目的として、UML 文書の作成支援するシステムの研究がすすめられているが、これも UML 文書の自動翻訳機能の提供にとどまっている。このような単純な翻訳機能の提供では、言語の背景にある文化やユーザの知識レベルの違いによって、交換・共有される情報の認識に関するギャップ(セマンティックギャップ)が生じることがある。

本研究は、共有対象のコンテンツそのものの翻訳に加えて、その意味を記述したメタデータを個々のユーザの言語・知識レベルに応じて変換して提示しようというものである(図1)。

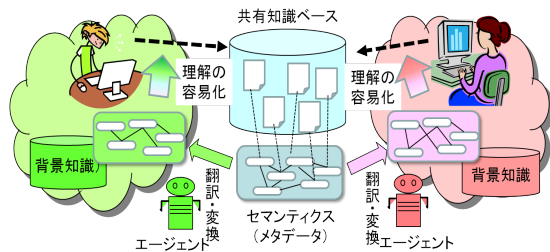


図1 研究の概要

2. 研究の目的

本研究では、メタデータを対象とする、オントロジーマッピングに基づく翻訳・変換を行う、という2つの基本的なアイデア(図2参照)に基づき、以下の3項目に関する技術を確立することを目的とする。

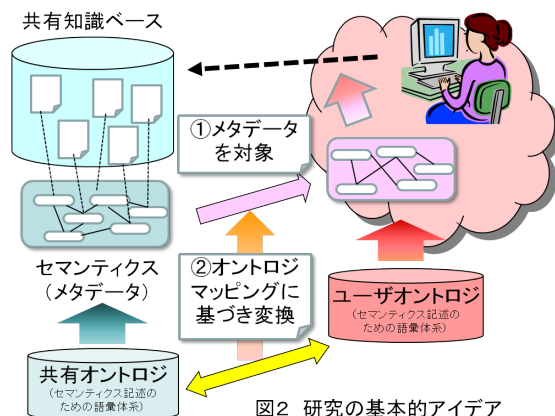


図2 研究の基本的アイデア

(1) メタデータ翻訳

RDF (Resource Description Framework) で記述されたメタデータの翻訳する手法を明らかにする。ここでは、RDF に用いられる語彙体系であるオントロジーを辞書として

用いる翻訳手法を検討する。

(2) オントロジーマッピング

各ユーザの持つ知識体系(オントロジー)は、言語も異なり、また同じドメインに関するものでも必ずしも同一ではない。そこで、共有知識ベースのオントロジーと各ユーザのオントロジーとの対応(マッピング)をとり、意味的に同一の部分を検出すると同時に、差異を検出する方法を開発する。この対応関係をもとに、オントロジーが異なる場合においてもメタデータの翻訳を可能とする方法を明らかにする。

(3) メタデータ変換

メタデータをそのまま翻訳しても、ベースとなるオントロジーが異なる場合にはユーザは理解することができない。そこで、オントロジー間の対応関係をもとに、共有知識ベースのメタデータを解釈して各ユーザが理解可能な形に変換するアルゴリズムを確立する。オントロジー間の差分をもとに、必要に応じてメタデータを補完する情報を生成する方法を検討する。

3. 研究の方法

具体的な研究の題材として、異文化間での表現の差が大きい料理レシピを題材とし、メタデータやオントロジーを構築して実証的に研究を行う。

(1) メタデータおよびオントロジーの構築

メタデータの表現には、セマンティック Web で標準的に用いられる RDF (Resource Description Framework) を用いる。またオントロジーの記述には、標準化が進められている OWL (Web Ontology Language) を用いる。

実用面を考慮し、メタデータ変換の前段として、Web 上に多く存在する料理レシピサイトからメタデータを構築する手法について研究を行う。またオントロジー構築に関しては、Wikipedia を情報源としたオントロジー構築方法について検討を行う。

(2) メタデータ翻訳

異なる言語の RDF で記述されたメタデータの翻訳を行うために、それぞれの RDF 記述のベースとなる異言語オントロジー間の概念の対応関係をもとに翻訳を行う手法を検討する。

(3) オントロジーマッピング

セマンティクス記述のための語彙体系であるオントロジーは、同一分野に関するものであっても、ユーザのバックグラウンドや知識レベルによって必ずしも同じであるとは限らない。セマンティクス(メタデータ)の変換のためには、異なる言語、異なる文化背景のもとに構築されたオントロジー間のマッピングを行い、その差異を検出する必要がある。さらに、語彙が欠けている部分については、対応するオントロジーから自動的に生

成する方法を開発する。

4. 研究成果

(1) メタデータ構築とオントロジー構築

国内外の料理レシピサイトを調査した結果を基に、構築するメタデータの RDF スキーマを決定した。主なクラスとしては、タイトル、作者、材料、料理手順がある。

このうち、タイトル、作者、材料等はレシピサイトの HTML を解析することで抽出可能であるが、調理手順は自然言語で記述されていることから、その解析が必要である(図3参照)。

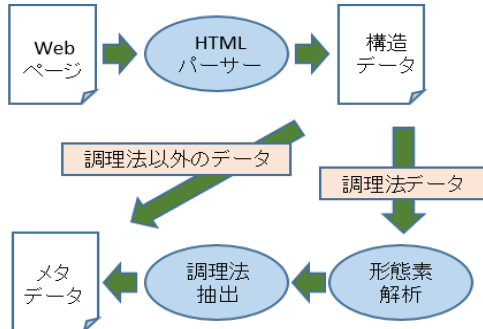


図3 料理レシピのメタデータ構築手順

調理手順の解析では、形態素解析を行った後、調理動作を表す動詞とそれに対応する調理対象となる材料名刺を探索することにより調理手順を生成する(図4)。

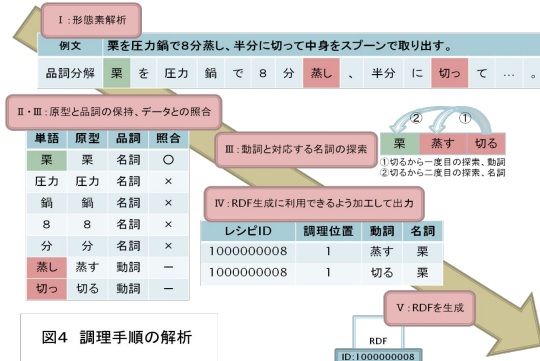


図4 調理手順の解析

オントロジー構築に関しては、Web 上の百科事典である Wikipedia を利用する方法を検討した。Wikipedia のカテゴリーは Wikipedia の文書の階層関係を表現しており、文書のタイトルを概念と考えれば、一種のオントロジーであるとみなすことができる。しかし、Wikipedia カテゴリーはアドホックに作成されていることから、特に上位概念に関しては整理されておらず、オントロジーとして利用するには無理がある。そこで、上位概念に関しては、既存の Word Net や日本語語彙体系など手動で構築された良質のオントロジーを利用し、これと Wikipedia カテゴリーを連結することを考えた。しかし、単純に概念の名称でマッチングをとると、同名異義語の問題が生じることから、Wikipedia カテゴリーの文書から抜き出したキーワードと、

既存オントロジーの概念の上位概念とのマッチングをとることにより正しい連結を実現する方法を考案した。

(2) メタデータ翻訳

メタデータの翻訳は、テキストの翻訳とは異なる。メタデータは元のデータに表れる概念とその間の関係で表現され、言語表現は単なる概念の表記方法に過ぎない。概念はその言語に対応するオントロジー中へのリソース参照で表現される。したがって、メタデータを翻訳するという事は、そのベースとなるオントロジーを翻訳対象言語のオントロジーに切り替えることに相当する(図4)。

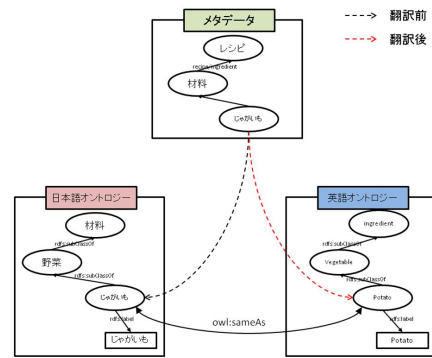


図5 メタデータ翻訳の概念

これを実現するためには、異なるオントロジー間で同じ概念を対応付けるためのオントロジーマッピングが必要となる。

(3) オントロジーマッピング

異言語のオントロジー間でのマッピングを実現するため、本研究では多言語で構築されている Wikipedia を利用する。最も簡便な方法は、Wikipedia の他言語リンクを利用する方法であるが、必ずしも他言語リンクが存在するとは限らない。その場合に対応する Wikipedia 文書を発見する方法として、以下の手順による方法を提案した(図6)。

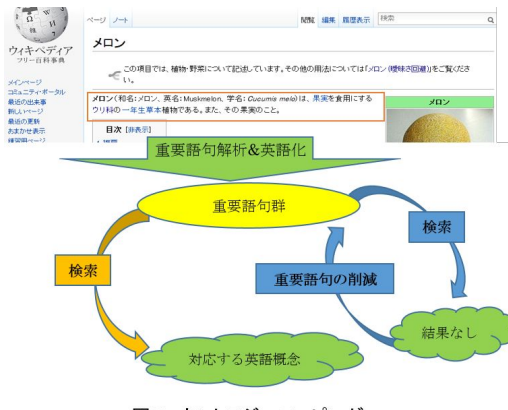


図6 オントロジーマッピング

対応づけようとする概念の名称をタイトルに持つ Wikipedia 文書を検索し、その概要説明テキストからキーワードを抽出する。TF-IDF 手法により、キーワードの重要度

を算出する。

キーワードを辞書によりターゲットの言語に音訳する。

翻訳したキーワードを用いて、ターゲットの言語の Wikipedia 文書を検索する。その際にキーワードの重要度を考慮する。検索された Wikipedia 文書のタイトルを対応する概念とする。

実験の結果、マッピング自体は 95%以上で成功するが、マッピング精度は 70%程度にとどまっており、キーワードの重みづけとそれに基づく Wikipedia 文書の検索方法に改良が必要である。

もう一つの問題として、文化の違い等により対応する概念が存在しない場合がある。その場合には、上位概念の検索を行い、上位概念間でマッピングを行った後、上位概念とのマッピング関係を構築する(図7)。

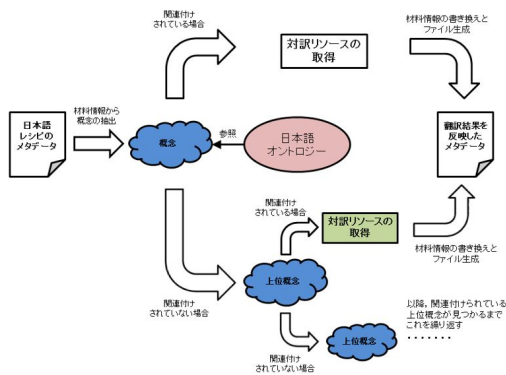


図7 対応する概念が存在しない場合のマッピング

(4) 結論と今後の展望

異文化コラボレーションにおけるセマンティックギャップを解消するという目標に向け、セマンティクスが明確なメタデータの翻訳方法について、基本的な手法を明らかにした。また、その実現のキーとなるオントロジーマッピングについて、Wikipedia を活用した実用的な方法を提案し、プロトタイプを構築することによってその実現性を明らかにした。

今後の課題として、オントロジーマッピングの精度の向上とともに、料理レシピ以外の多様な応用に向けた検討が必要である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

著者名：奥健太、服部文夫、論文標題：セレンディピティ指向情報推薦のためのフュージョンベース推薦システム、雑誌名：知能と情報、査読：有、巻：25、発行年：2013、ページ：524-539。

著者名：Homan, T., Morita, M., Yamaguchi, Y., Kuwabara, K.、論文標題：Construction of a Topic Database

for Remote Conversation Support、雑誌名：Int. Journal. Of Software Science and Computational Intelligence、査読：有、巻：5、発行年：2013、ページ：77-91。
著者名：Yasuda, K., Kuwahara, N., Kuwabara, K., Morimoto, K., Tetsutani, N.、論文標題：Daily Assistance for Individuals with Dementia via Videophone、雑誌名：American Journal of Alzheimer's Disease and Other Dementias、査読：有、巻：28、発行年：2013、ページ：508-516。
著者名：Morita, M., Kuwabara, K.、論文標題：Agent-based customization of a remote conversation support system、雑誌名：Intelligent Decision Technologies、査読：有、巻：7、No.3、pp.175-184、2013。
著者名：奥健太、西崎剛司、服部文夫、論文標題：地域限定性スコアに基づく位置情報付きコンテンツからの地域限定語句の抽出、雑誌名：情報処理学会論文誌(データベース) 査読：有、巻：5、発行年：2012、ページ：97-116。

[学会発表](計22件)

著者名：Kuwabara, K. Ohta, N.、発表標題：Toward a Crowdsourcing Platform for Knowledge Base Construction、学会等名：Sixth International Conference on Information, Process, and Knowledge Management、発表年月日：2014年3月25日、発表場所：Barcelona(スペイン)。
著者名：Kinomura, S., and Kuwabara, K.、発表標題：Developing a Multilingual Application Using Linked Data: A Case Study、学会等名：5th International Conference on Computational Collective Intelligence、発表年月日：2013年9月12日、発表場所：Craiova(ルーマニア)。

発表者名：Nilar Aye, Hlaing Su Khin, Toe Toe Win, Tayzar, KoKo, MoMo Zin Than, Fumio Hattori, Kazuhiro Kuwabara、発表標題：Multi-Domain Public Key Infrastructure for Information Security with Use of a Multi-Agent System、学会等名：5th Asian Conference on Intelligent Information and Database Systems、発表年月日：2013年3月19日、発表場所：Kuala Lumpur(マレーシア)。
著者名：Homan, T., Morita, M., Yamaguchi, Y., and Kuwabara, K.、発表標題：Topic Database for Conversation Support: Its Application and Access Control、学会等名：11th IEEE International Conference on Cognitive

Informatics & Cognitive Computing、発表年月日：2012年8月22日、発表場所：立命館大学（京都府）
著者名：Kuwabara, K., and Morita, M.,
発表標題：Dynamic Customization of a Remote Conversation Support System: Agent-Based Approach、学会等名：6th International KES Symposium on Agents and Multi-agent Systems、発表年月日：2012年6月27日、発表場所：Dubrovnik（クロアチア）
著者名：Oku, K. Hattori F.、発表標題：Fusion-based Recommender System for Improving Serendipity、学会等名：International Workshop on Novelty and Diversity in Recommender Systems、発表年月日：2011年10月23日、発表場所：Chicago（米国）
著者名：Kuwabara, K. and Yasunaga, S.、発表標題：Use of Metadata for Access Control and Version Management in RDF Database、学会等名：Knowledge-Based and Intelligent Information and Engineering Systems
発表年月日：2011年9月13日、発表場所：Kaiserslautern（ドイツ）

6. 研究組織

(1) 研究代表者

服部 文夫 (HATTORI FUMIO)
立命館大学・情報理工学部・教授
研究者番号：80388131

(2) 研究分担者

桑原 和宏 (KUWABARA KAZUHIRO)
立命館大学・情報理工学部・教授
研究者番号：10374092