

日本語 Wikipedia オントロジーの
自動構築と評価

2013 年度

玉川 奨

主 論 文 要 旨

報告番号	甲 乙 第	号	氏 名	玉川 奨
主 論 文 題 目： 日本語 Wikipedia オントロジーの自動構築と評価				
(内容の要旨) 大規模オントロジーは、データ統合などの情報基盤として期待されているが、オントロジーの自動構築には、コストと保守に大きな課題を抱えている。その課題を解決するために、フリーテキストからのオントロジー自動構築が試みられてきたが、自然言語理解に限界があり、実用レベルに到達しないことから、近年、半構造情報を有する情報資源からオントロジーを自動的に構築する方法が注目されてきた。以上の背景から、本論文では、Web 上のオンライン百科事典である日本語版 Wikipedia の有する半構造情報から、インスタンスの抽出、概念抽出、概念間の上位下位関係、インスタンス間の意味的關係（プロパティ）、プロパティの定義域と値域、プロパティタイプなど、オントロジーとして重要な情報を抽出する方法を検討し、大規模汎用オントロジー（日本語 Wikipedia オントロジーと呼ぶ）を自動構築し、日本語 Wikipedia オントロジーの領域オントロジー構築支援可能性と Linked Open Data のハブの観点から、その有用性を評価する。 以下に本論文の構成を示す。 はじめに、第 1 章において、本研究の背景、問題、目的について述べる。 第 2 章では、本研究の関連技術として、オントロジーの定義および具体例、オントロジー構築方法論、Wikipedia、Linked Open Data について述べると共に、それらの関連研究についても述べる。 第 3 章では、日本語版 Wikipedia から概念および概念間の関係（is-a 関係、クラスーインスタンス関係、プロパティ定義域、プロパティ値域、プロパティ上位下位関係、インスタンス間関係、その他の関係）を抽出することで、日本語 Wikipedia オントロジーを自動構築する手法の提案と各手法の評価について述べる。 第 4 章では、日本語 Wikipedia オントロジーの領域オントロジー構築支援としての評価について述べる。また、Linked Open Data としての設計と公開、Linked Open Vocabularies との連携による日本語語彙構築手法の提案と評価、検索支援ツール WiLD の設計と評価により、Linked Open Data のハブとしての評価について述べる。これらの評価から日本語 Wikipedia オントロジーの有用性を示す。 最後に第 5 章では、本論文のまとめと今後の課題および展望について述べる。				

SUMMARY OF Ph.D. DISSERTATION

School Open and Environmental Systems	Student Identification Number	SURNAME, First name TAMAGAWA Susumu
Title Building up Japanese Wikipedia Ontology with Semistructured Information		
Abstract <p>Large-scale ontologies are expected to work as an information infrastructure for information services, such as information retrieval and data integration. Because it takes many costs for human experts to build and maintain ontologies, much attention has been come to the work on automatic ontology construction from free text. However, natural language processing still has much limitation to free text and so the work has not been in practice yet. Thus more attention moves to automatic ontology construction from semi-structured information resources, such as Wikipedia.</p> <p>This dissertation discusses how to extract important information to compose ontologies from Japanese Wikipedia (Japanese Wikipedia Ontology). They include instances, classes, super-sub relationships between classes, properties between instances, property domains and ranges, and property types. Furthermore, Japanese Wikipedia Ontology has been evaluated from the following points: how much to support for human experts to build up domain ontologies and how much it works as Japanese Linked Open Data Cloud Hub.</p> <p>This dissertation has the following structure.</p> <p>Chapter 1 describes backgrounds and goals of this research.</p> <p>Chapter 2 explains what ontologies in information science are, and shows us ontology development process and environment, Wikipedia, and Linked Open Data.</p> <p>Chapter 3 discusses how to extract important information to compose Japanese Wikipedia Ontology methods with extraction metrics, such as precision.</p> <p>Chapter 4 evaluates how much Japanese Wikipedia Ontology support human experts to build up domain ontologies in the field of hydroelectricity and how much Japanese Wikipedia Ontology works as Japanese Linked Open Data Cloud Hub.</p> <p>Chapter 5 wraps up lessons learned from building Japanese Wikipedia Ontology with case studies and shows us what future issues are.</p>		