

Vocabulary Building Mechanism based on Lexicographic Semantics for Querying the Semantic Web

(セマンティックWeb に対する問合せのための辞書的意味論に基づく語彙構築機構)

学位論文内容の要旨

Web リソースは、文書や画像、音楽、動画にとどまらず、ネットワークで繋がれたモバイル機器や電化製品、環境の中に埋め込まれたコンピュータやセンサーなども対象にする動きがある。今後もこの動きが加速し、様々なものが Web を介してアクセス可能になり、今以上に多くのリソースが Web 上に存在することが予想される。そこで、それらの膨大な Web リソースを効率的に利用するための方法として、Web リソースの意味内容情報を用いた検索を可能とするセマンティック Web の実現がある。そのための研究は現在も盛んにおこなわれており、World Wide Web Consortium (W3C) を中心に国際的な標準化が進められている。その標準には、Web リソースの意味内容情報の記述のための構文である Resource Definition Framework, RDF の記述に用いられる語彙を定義し、意味内容を記述するための RDF-schema, OWL web ontology language, リソース間の推論規則を記述するための Semantic Web Rule Language, 問合せのための SPARQL Protocol and RDF Query Language (SPARQL) がある。他にも、Web リソース間の関係記述のための語彙 Dublin Core や人の関係記述のための記述語彙 FOAF など意味内容情報の記述のための語彙が定義されている。また、定義した語彙を用いて Wikipedia のデータを構造化した DBpedia やコンピュータサイエンス分野の学術的出版物のデータを構造化した DBLP Bibliography があり、両者とも SPARQL を用いた問合せも可能である。また、これらのデータを繋ごうとする Linked Data という試みもあり、膨大な量のリソースから成るセマンティック Web 実現の動きがある。セマンティック Web 技術が有用されている例には、Rich Site Summary を用いたニュースサイトや Web ログがあり、タイトルや更新日時、要約などの情報をリソースに付加することで内容情報による検索を可能にし、更新日時が新しい記事や要約にあるキーワードを含む記事の収集を機械処理により自動化できる。

意味内容に基づいた Web リソース検索が可能となる環境が整いつつある一方で、検索に用いられる W3C が提案する提案問合せ言語 SPARQL の使用には、クラスやプロパティ、変数の知識が必須であり、それらに関する知識のない一般的ユーザによる使用は困難である。

本研究では、セマンティック Web に対して、拡張可能で再利用性の高い辞書的意味論に基づく語彙(辞書の語彙)を定義し、その語彙を用いた問合せ言語とその評価機構を提案することにより、上記の問題を解決するための理論的基盤となる研究を行った。本研究における成果は以下の2点である。

1. セマンティック Web に対する問合せのための辞書的意味論に基づく語彙構築機構の提案
2. 1の機構によって定義された語彙を用いた問合せ言語の提案

本論文の構成は以下の通りである。第1章では、本研究の背景や動機、貢献について記す。第2章では、セマンティック Web における Web リソースの意味内容記述のための語彙の定義言語と、セマンティック Web に対する問合せ言語の研究について解説するとともに、それらにおける問題点を提起する。第3章では、本研究の基盤となるセマンティック Web に基づいた辞書的意味論について解

説し、辞書的語彙の基となる基本語彙の定義について示す。基本語彙は、セマンティック Web において定義されているプロパティやクラスにマッピングして定義される。

第 4 章では、辞書の意味論に基づいて定義した辞書的語定義言語 Lexicographic Word Definition Language (LWDL) を用いた語彙構築機構を提案する。この語彙構築機構により、第 3 章で定義した基本語彙を基に、派生語彙を定義することで、語彙を拡張することができる。提案手法では、既定義語の組み合わせに新たなラベルを与えることで、その新たなラベルを新たな派生語として定義することができる。例えば、「Parent」と「Male」という語が存在する時、それらの積の組み合わせ「Parent and Male」に対して新たなラベル「Father」を与え、新たな語「father」を定義することができる。このような語彙の拡張により、「Parent and Male」という語の組み合わせ表現の代わりに、直観的で短い表現「Father」を用いて表現することができる。また、語彙の要素には、自然言語における名詞と形容詞と動詞があり、これらを組み合わせて複雑な対象を表現することができる。動詞は、状態の記述と状態の変更の 2 つを表すために用いられることが考えられるが、本論文では前者を表すために用いることとする。

第 5 章では、辞書的語彙を用いた問合せ言語 Query Language based on Lexicographic Semantics (QLLS) の構文を示す。第 4 章で定義した辞書的語彙に含まれる名詞と形容詞と動詞を用いた語の組み合わせにより複雑な対象を表現することができる。第 6 章では、第 5 章で提案する問合せ言語 QLLS での問合せの評価機構について示す。提案評価機構は複数の書き換え規則から成る。それらの書き換え規則を適切な順に適用することで、第 1 に QLLS 表現は関係演算式に書き換えられ、次に関係演算式は対応する SPARQL 表現に書き換えられる。

第 7 章では、第 4 章～第 6 章で示した提案手法を組み合わせたシステムを用いて、セマンティック Web に対して辞書的語彙を定義し、その語彙を用いた問合せと、その問合せの評価方法について例を用いて示す。第 8 章では関連研究との比較について述べ、第 9 章に結論を示す。

本論文では、拡張可能で再利用性の高い辞書的意味論に基づく語彙をセマンティック Web に対する問合せ言語に導入する方法を提案することで、語と語の組み合わせによる簡便な表現による問合せを実現した。また、一度定義した提案語彙は、限られた数の基本語彙のみを問合せ対象のセマンティック Web にマッピングすることで、マッピングしたセマンティック Web に対する問合せに再利用することができる。

学位論文審査の要旨

主 査 教 授 田 中 譲
副 査 教 授 原 口 誠
副 査 教 授 有 村 博 紀

学 位 論 文 題 名

Vocabulary Building Mechanism based on Lexicographic Semantics for Querying the Semantic Web

(セマンティックWeb に対する問合せのための辞書的意味論に基づく語彙構築機構)

近年、膨大な Web リソースを効率的に利用するための方法として、Web リソースの意味内容情報を用いた検索を可能とするセマンティック Web が提案され、World Wide Web Consortium (W3C) を中心に国際的な標準化が進められている。Web リソースの意味内容情報の記述のための構文である Resource Definition Framework(RDF) に用いられる語彙を定義し意味内容を記述するための RDF-schema と OWL web ontology language, リソース間の推論規則を記述するための Semantic Web Rule Language, 問合せのための SPARQL Protocol and RDF Query Language (SPARQL) が提案され普及しつつある。さらに、定義した語彙を用いて Wikipedia のデータを構造化した DBpedia や、コンピュータサイエンス分野の学術的出版物のデータを構造化した DBLP Bibliography が既に構築され公開されており、両者とも SPARQL を用いて問合せ可能になっている。

意味内容に基づいた Web リソース検索が可能となる環境が整いつつある一方で、検索に用いる問合せ言語として W3C が提案している提案問合せ言語 SPARQL を活用するには、クラスやプロパティ、変数の知識が必須であり、それらに関する知識のない一般ユーザによる使用は困難である。

本研究では、セマンティック Web に対して、拡張可能で再利用性の高い辞書的意味論に基づく語彙(辞書の語彙)を定義し、その語彙を用いた問合せ言語とその評価機構を提案し、それらの意味論の理論的基盤を構築し、実働可能なシステムの開発基盤を明確にしている。本研究における成果は以下の2点である。

1. セマンティック Web に対する問合せのための辞書的意味論に基づく語彙構築機構の提案と、その意味論の理論的基盤の構築。
2. 上記の機構によって定義された語彙を用いることが可能な問合せ言語の提案と、上記の理論的意味論に基づく評価機構の構築。

本論文の構成は以下の通りである。第1章では、本研究の背景や動機、貢献について記している。第2章では、セマンティック Web における Web リソースの意味内容記述のための語彙定義言語と、セマンティック Web に対する問合せ言語に関する従来の研究について解説するとともに、それらにおける問題点を提起している。第3章では、本研究の基盤となるセマンティック Web に基づいた辞書的意味論の理論について解説し、辞書の語彙の基となる基本語彙の定義法について示している。第4章では、辞書的意味論に基づき、辞書の語定義言語を用いた語彙構築機構を提案している。この語彙構築機構により、第3章で定義した基本語彙を基に、派生語彙を定義することができ、これにより語彙を拡張することができる。語彙の要素には、自然言語における名詞と形容詞と動詞の別があり、

これらを組み合わせて複雑な概念を表現することができる。動詞には、状態の記述を表す場合と状態の更新を表す場合があるが、本論文では前者の場合のみが考察の対象となっている。第5章では、辞書的語彙を用いた問合せ言語の構文が示されている。第4章で定義した辞書的語彙に含まれる名詞と形容詞と動詞を用いた語の組み合わせにより複雑な対象を表現することができる。第6章では、第5章で提案された問合せ言語 QLLS で記述された問合せの評価機構が示されている。評価機構は複数の書き換え規則から成り、それらを適切な順に適用することで、QLLS 表現は最初に関係演算式に書き換えられ、次に対応する SPARQL 表現に書き換えられる。第7章では、第4章～第6章で示した提案手法を組み合わせて開発されたシステムを用いて、セマンティック Web に対して先ず辞書的語彙を実際に定義し、その語彙を用いた問合せと、その問合せの評価過程について例に即して説明がなされている。第8章では関連研究との比較評価が行われ、本研究の優位性が示されている。第9章には結論が示されている。

これを要するに、著者は、近年ますます普及しつつあるセマンティック Web に対する使い易く簡便で柔軟な検索問い合わせ記述を実現することを目指して、(1) セマンティック Web に対する問合せのための辞書的意味論に基づく語彙構築機構と、その意味論の理論的基盤、(2) このようにして定義された語彙を用いることができる問合せ言語と、理論的意味論に基づくその評価機構に関する理論的基盤とシステム実現法に関して新知見を得たものであり、ウェブ工学ならびに知識工学に対して貢献するところ大なるものがある。よって著者は、北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。