オントロジー更新時における検索障害の分析と回避手法

柳田 憲士郎† 塚本 享治†
東京工科大学大学院バイオ情報メディア研究科

1 はじめに

近年、インターネット上のドキュメントに対して意味情報を用いた検索手法としてセマンティックWEB技術が提唱されている。セマンティックWEB技術では、OWLによってメタデータ化された意味情報（オントロジー）を用いることにより、意味検索を可能としている。しかし、検索システムで利用されているオントロジーをオントロジー提供者が自由に編集した際に推論結果の変化により、検索システムが検索不能に陥る可能性がある。

本報告では、オントロジー提供者がどのような編集を行った際に検索システム側が検索不能に陥るかを分類し、検索不能に陥らないオントロジーの構築及び回避手法を提案する。

2 背景

セマンティックWEB技術では、OWLを用いることにより、[1]といったe-ラーニングや[2][3]のように専門的知識を必要とする分野における情報の分類及び検索を行う手法として用いられ、研究がなされている。

OWLでは、RDFのデータ構造により規格上は新しいデータや推論記述の追加・変更が既存のデータ部分を変更せずに行うことができ、またWEB上に公開された別のオントロジーを自分のオントロジー構築に利用することができる。

しかし、これらの多くは、オントロジーの追加・変更を考慮されたものではなく、変更内容がOWLの構文にそったものであっても推論記述によっては、オントロジーを読み取って検索を行っているサービスの検索能力低下及び検索不能に陥る可能性がある。これは、WEB上に公開されたオントロジーがどのような検索サービス及び別のオントロジー構築に利用されているかということをオントロジー提供者が知ることができないからである。

本研究では、どのようなオントロジー構築・編集作業が検索障害を発生させるか、オントロジー更新状況の分類を行う。また、オントロジー提供者が公開中のオントロジー利用状況を知ることができない環境であっても、オントロジー利用側に与える影響を最小限に抑えるためのオントロジー構築及び回避手法の提案を行う。

3 オントロジー構築及び利用条件

本研究では、表1の条件化で構築されたオントロジー及び検索サービスについて、検索障害の分析及び回避手法の提案を行う。OWL規格については、記述論理性、計算処理効率、対応しているライブラリ及びソフトウェアを考慮し、W3Cによって定義されているOWL2.0[4]を用いる。

表1：オントロジー構築及び検索環境

オントロジー記述言語	OWL2.0 DL
オントロジー編集ソフト	Protégé 4.0
推論エンジン	Pellet 2.0 RC4
検索クエリ言語	SPARQL,ARQ
クエリ実行ライブラリ	Jena 2.5

オントロジー構築及び推論エンジンにはOWL2.0の推論記述に対応しているProtege4.0、Pellet2.0RC4を用いる。これらのオントロジーはW3Cによって標準化されているRDF検索クエリ言語SPARQL及びその拡張言語であるARQ[5]を用いる。これらクエリ言語処理にはセマンティックWEBアプリケーションの開発を目的として作られたライブラリであるJena2.5を用いた。

4 検索障害を発生させる更新作業の分析

本研究では、[6][7]で行ったオントロジー構築時において発生した検索障害及び更新作業について表2のように分類した。

表2：検索障害を発生させる更新作業の分類

検索障害内容	原因となった更新作業
検索対象到達不能	・URIの変更 ・クラス、プロパティの細分化
矛盾表記による推論不能	・クラス、プロパティへのEquivalent指定
検索処理増加に伴うシステムダウン	・オブジェクトプロパティに対する推移性指定 ・Owl:diffrentFromタグの付与
想定外の検索結果変更	・推論記述変更による不必要なデータ抽出

検索対象到達不能は、いままで検索サービスが検索可能であったものが、オントロジー更新作業によって、検索対象としてはずされ、検索結果として取得できないという問題である。クラス、プロパティ名の変更や細分化といった、URIの変更によるものが主である。

矛盾表記による推論不能は、二つの相反する推論記述によって推論矛盾を引き起こし、推論エンジンが推論不可能であるという結果を返すものである。これは、Equivalent指定により、クラスやプロパティが同一であるとみなす記述を行った際に、それぞれのクラス、プロパティが保有している推論記述が異なっているという理由で発生することが多い。

検索処理増加に伴うシステムダウンは、推論記述変更による推論結果の増大により、検索サービスが使用可能なリソースを超えてしまうことによりおきる。推論結果の爆発的増加を引き起こす推論記述の理由に、オブジェクトプロパティに対する推移性指定及びOwl:diffrentFromタグの付与が上げられる。

想定外の検索結果変更は、オントロジー提供者がより詳細にオントロジー記述を行おうとした結果、検索サービス側に不必要な検索結果が追加されるというものである。

Analysis and evasion technique of retrieval trouble when updating ontologies
Kenshiro YANAGIDA†, Michiharu TSUKAMOTO†
†Tokyo University of Technology Graduate School

5　回避手法の提案

検索障害及びその理由となった更新作業を分析した結果、公開中のオントロジーに対して、以下の更新制約を設けることにより、検索障害を回避できるのではないかと考えた。

(1)　URI 表記変更の禁止
(2)　細分化前のクラス及びプロパティを残し、細分化後のクラス、プロパティの親として記述する
(3)　論理性質の無いプロパティを論理性質のあるプロパティの子プロパティとして記述し、インスタンスの登録は子プロパティに対して行う
(4)　自分の管理下にないオントロジーが提供するクラスに対する値制限記述変更の禁止
(5)　推論表記のあるクラス、プロパティに対する equivalent 記述の禁止。intersectionOf, unionOf、subPropertyOf による代用
(6)　大量のインスタンスが利用しているプロパティに対する推移性記述の禁止及び propertyChain による代用

6　実験

本研究では図１で構成されるオントロジー及び検索環境下でのオントロジーに対する検索実験を行った。この検索実験は、大学内で公開されているシラバスについて、技術用語及び学生の履修状況といった異なるオントロジー間を結びつけ、推論処理を用いた検索が行うことが可能である。図１で示される３つのオントロジーに対して手法(1)～(6)を用いた場合とそうでない場合のオントロジー更新作業を行った。

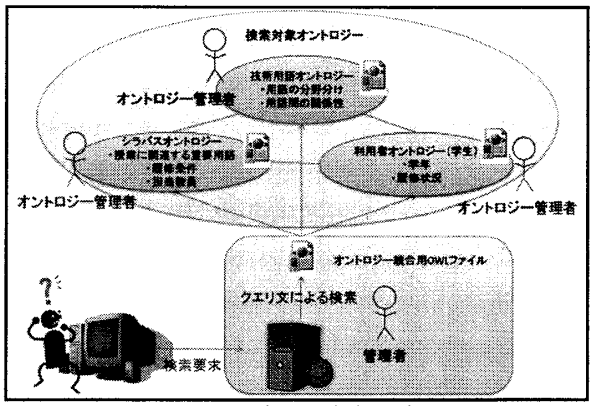

図１：オントロジーと検索環境の関係図

7　実験結果

実験では、格提案手法を用いる前後における検索障害の発生状況と提案手法を用いたことによる、検索速度及び検索結果の変化について確認を行った。表３は障害発生の再現性を表したものである。なお、表３に表記されている数値は以下の状況を表している。

０…再現性無し
１…再現性低い（再現率 1%以下）
２…再現性有り
３…再現性常時有り

表３：手法利用前後における障害発生変化

	提案手法利用前				
手法	到達不能	推論不能	処理増大	不必要抽出	合計値
(1)	3	0	0	0	3
(2)	3	0	0	0	3
(3)	2	0	0	2	4
(4)	0	2	0	0	2
(5)	0	3	2	0	5
(6)	0	2	3	2	7
	提案手法利用後				
手法	到達不能	推論不能	処理増大	不必要抽出	合計値
(1)	0	0	0	0	0
(2)	0	0	0	0	0
(3)	0	0	2	0	2
(4)	0	0	0	0	0
(5)	0	0	0	0	0
(6)	1	0	0	0	1

8　考察

実験から提案手法を用いることにより、推論不能障害は回避可能であるという結果が得られた。

到達不能障害については、提案手法を用いることによりほぼ解消することができた。(6)の手法により、到達不能障害の可能性が発生したのは、推移性記述によって得られるデータが検索サービスに必要であったためである。これについては、オントロジー統合用 OWL ファイル側に推移性記述を移すことにより、対策が可能であると考える。

処理増大に関する障害では、システムダウンが発生する障害については取り除くことができたが、(3)の手法を用いた際に、大量のインスタンスが所属するプロパティであった場合に、トリプル量の増加量が大きいために、処理が増大する恐れが発生した。

9　結論

検索サービスにすでに利用中の状況下にあるオントロジーに対する更新作業によって発生する障害及び発生理由を分析し、障害を回避する手法の提案を行った。提案手法を用いることにより、多くの検索障害を回避することができた。

今後はオントロジー検索環境の実験対象を増やし、より多くの検索障害発生状況を考慮した回避手法の提案を行いたい。

参考文献

[1]高橋圭仁, 安孫子友美, 根岸英梨子, 板橋吾一, 加藤靖, 高橋薫: "オントロジーに基づいた暗号学習用の e-Learning システム," 電子情報通信学会技術研究報告, ET2004-73, pp.1-6 (2004).
[2]国府裕子, 周俊, 古崎晃司, 今井健, 大江和彦, 溝口理一郎: 臨床医療オントロジーの構築に関する基礎的な考察, 人工知能学会第 22 回全国大会(JSAI2008), 2E3-01
[3]山田篤, 安達文夫, 海田茂, 今門政記, 河合正樹, 小町祐史: 博物館情報の知的横断検索のためのフレームワーク, 画像電子学会 VMA 研究会博物館・美術館 DTD-SG
[4]W3C-"OWL 2 Web Ontology Language", http://www.w3.org/TR/owl2-semantics/,2008
[5]ARQ - A SPARQL Processor for Jena, http://jena.sourceforge.net/ARQ/,2004
[6]柳田憲士郎, 塚本享治:大学内情報オントロジーの構築と検索システムの実装 情報処理学会第 69 回全国大会
[7]柳田憲士郎, 塚本享治:セマンティック WEB 技術を用いた技術ドキュメントの類似性検出方法とその評価, FIT2008 第 7 回情報科学技術フォーラム