

# 利用者からのフィードバック情報を用いたオントロジ拡充技術

境美樹 佐藤宏之 中村昌志

日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

## 概要

オントロジを利用したポータルサービスを提供する際、利用者のニーズや流行によりオントロジが変化していくことが望まれる。本報告では、利用者の口コミ情報などのフィードバック情報を用いたオントロジ拡充技術による解決方式を提案する。更に本方式をインターネットタウンページのデータを使用して評価した結果を報告する。

### 1. はじめに

Web 上には、goo[1]やインターネットタウンページ[2]などの検索ポータルを始めとした様々なポータルサービスが存在する。これらは、新しい情報や必要な情報を網羅的かつ効果的に閲覧したいという利用者のニーズに応える事ができる。

スタンフォード大学の TAP プロジェクト[3]では、Semantic Search/ Activity Based Search という検索サービスのデモを公開している。TAP KB という一般的な概念に対する語彙や分類情報を含む知識ベースを利用しており、入力キーワードに対するテキスト検索の結果だけでなく関連する情報も表示できる。このように、ポータルサービスに語彙の階層や制約などを記述したオントロジを利用する事で、利用者のニーズに沿った情報の提供が行える。

一般的にオントロジは普遍的なものである。一方、ポータルサービスでは日々新しい情報が提供されており、利用者のニーズは変化していく。そのため、ポータルサービスに利用するオントロジは利用者や社会の状況に合わせて変化していくことが望まれている。

本報告では、利用者の口コミ情報などのフィードバック情報を用いたオントロジ拡充技術を提案する。

以下、2章ではオントロジを用いたポータルサービスとその問題点について述べ、3章ではその解決方法を述べる。4章では、インターネットタウンページのデータを利用して実装したポータルサービスを紹介し、3章で述べた解決方法の適用を考える。そして、5章で評価を行い、6章では評価結果に対する考察を行う。最後に7章で今後の課題などを述べる。

### 2. オントロジを用いたポータルサービスの問題点

#### 2.1 オントロジを用いたポータルサービス

オントロジを用いたポータルサービスを提供する場合、サービスのドメインに応じたオントロジを準備する必要がある。例えば、ラーメンに関するポータルサービスの場合、図1の様な Resource Description Framework (RDF) [4]と RDF Vocabulary Description Language 1.0: RDF Schema (RDF スキーマ) [5]を用いてグラフ表現したラーメンに関するオントロジが考えられる。「ラーメン」という概念から様々なラーメンの種類やラーメンを提供している業種を知る事ができる。

また、図1で記述されているメタデータは、リソース (サブジェクト)、プロパティ、プロパティ値の組から成る RDF を用いている。

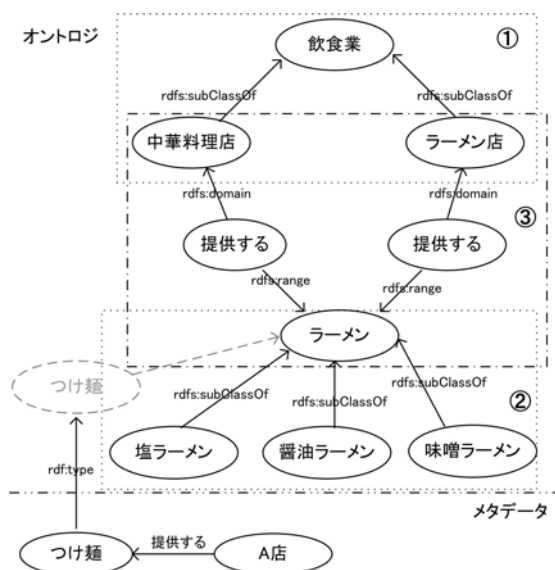


図1 ラーメンに関するオントロジとメタデータ

## 2.2 問題点

提供されたオントロジが不完全なものである場合や時間が経過して内容が古くなった場合、利用者の利便性を損なう。例えば、図1を見ると、A店ではつけ麺を扱っているのだが、オントロジとして定義されていない。そのため、利用者は「ラーメン」というキーワードからつけ麺を扱っている「A店」を得る事ができない。この問題を解決する手段の一つとして、管理者などが店舗に関する情報を得る度にオントロジを追加していく方法が考えられる。しかし、店舗から提供される情報や利用者から提供される情報が増加すると人手による構築や追加には限界があると考えられる。

## 3. 利用者フィードバック情報の利用

2章で述べた問題点に対する解決法として、ポータルサービスの利用者からフィードバック情報を収集し、オントロジ拡充に利用する方式を提案する。

提案方式では、以下の情報が予め準備されているものとする。

＜オントロジ＞

- ① リソースのクラス定義 (図1の①)  
例) <ラーメン店> <rdfs:subClassOf> <飲食業>
- ② プロパティ値として用いるクラス定義 (図1の②)  
例) <塩ラーメン> <rdfs:subClassOf> <ラーメン>
- ③ プロパティ制約定義 (図1の③)  
例) <提供する> <rdfs:domain> <ラーメン店>  
<提供する> <rdfs:range> <ラーメン>

＜メタデータ＞

### ・リソースの情報

- 例) <A店> <rdf:type> <ラーメン店>
- <A店> <電話番号> “03-xxxx-xxxx”
- <A店> <住所> “東京都・・・”

## 3.1 利用者フィードバック情報

ポータルサービス利用者のフィードバック情報

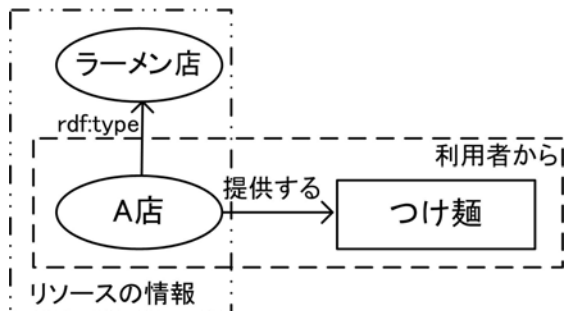


図2 利用者フィードバック情報

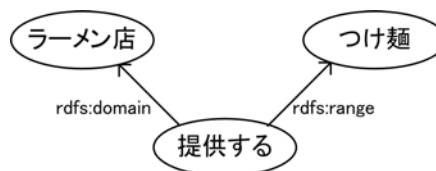


図3 生成されたプロパティ制約定義

報として以下が挙げられる。

- (1) 情報閲覧履歴
- (2) 検索履歴
- (3) あるリソースに関する口コミ的情報

(1)や(2)の情報からは、例えば、プロキシサーバを用いて利用者が閲覧した文書情報を取得し、その利用者にとって重要度の高いキーワードを抽出することは可能[6]だが、キーワード間の関係やプロパティとプロパティ値に関する情報を得る事ができない。そのため、今回は(3)の口コミ的なフィードバック情報、例えば2章で示した「A店では、つけ麺を提供している」といった利用者の経験に基づく情報をメタデータ化して利用する。

## 3.2 フィードバック情報を用いたオントロジ拡充方式

「A店では、つけ麺を提供している」というフィードバック情報のリソースは「A店」、プロパティは「提供する」、プロパティ値は「つけ麺」である(図2)。この時、リソースの情報として予め準備されているメタデータから、A店はrdf:typeとして「ラーメン店」を持つことが分かる。これを利用して、「提供する」というプロパティは、リソースとして「ラーメン店」のインスタンスを、プロパティ値として「つけ麺」を取ることがわかるのでプロパティ値と同名のクラスを生成して、図3に示す「提供する」に関するプロパティ制約定義を生成する。

しかし、入力されたプロパティ値が「A店特製ラーメン」の様な固有名詞だった場合、クラス及びプロパティ制約定義の生成は行わない。

## 3.3 テキスト解析とフィードバック情報を組み合わせたオントロジ拡充

3.2で提案した手法を活用して、更にオントロジを拡充する。ポータルサービスで紹介する情報(例えばラーメン店など)は、ホームページや掲示板、広告などのテキスト情報を持つことが多い。提案方式は、まず、このリソースについてのテキスト情報から既存のプロパティ制約定義を利用して、リソースに関するプロパティ値の抽出を行い、メタデータを生成する。定義されていない用語が抽出された場合、3.2の方式を用いて新たな

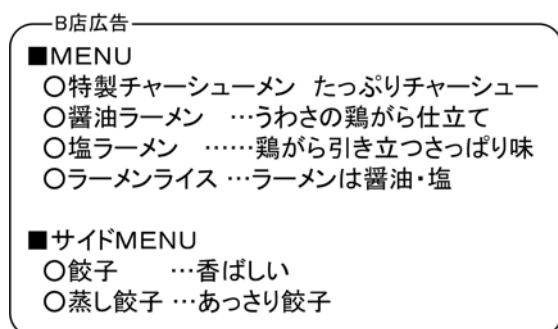


図 4 B店の広告

プロパティ制約定義を生成する。テキスト情報から抽出されたが、メタデータ生成に利用されなかったプロパティ値となりうる名詞などの用語は、プール情報としてストックする。プール情報は、その用語に関連するフィードバック情報が新たに入力された場合、そこからメタデータやプロパティ制約定義を生成するのに用いられる。

例えば、ラーメン店「B店」の広告（図4）を用いた場合を説明する。

まず、図1で示したオントロジを用いてラーメンに関するメタデータを抽出する。この時、広告データを形態素解析した結果から、名詞などを組み合わせた用語を取得する。そして、取得した用語と、プロパティ制約定義のプロパティ値として定義されている語彙の一部と一致している部分があればプロパティ値として抽出する。この結果、

〈B店〉〈提供する〉 “醤油ラーメン”

〈B店〉〈提供する〉 “塩ラーメン”

〈B店〉〈提供する〉 “ラーメンライス”

というメタデータが生成される。「ラーメンライス」はオントロジに存在しないので、「ラーメンライス」というクラスを生成し、「ラーメン店」が「ラーメンライス」を「提供する」というプロパティ制約定義をオントロジに追加する。

次にプール情報を抽出する。ここでは、上記で取得されたプロパティ値とテキスト上の文字の並びや直前の記号が同一など配置が類似している用語を抽出する。その結果として

〈B店〉 “特製チャーシューメン”

〈B店〉 “餃子”

〈B店〉 “蒸し餃子”

が抽出される。

ここで、利用者フィードバック情報として、「A店は餃子を提供する」という情報が入力されたとする。この情報から、3.2で述べた手法を用い、プロパティ制約定義を追加する。また、A店とB店は共にラーメン店であり、B店のプール情報に

「餃子」が存在する。この事実から、「B店は餃子を提供する」というメタデータが生成され、語彙の一部と一致していることから、「蒸し餃子」も取得され、「B店は蒸し餃子を提供する」というメタデータが生成される。「蒸し餃子」はまだオントロジに定義されていない語彙なので、3.2と同様の方式でプロパティ制約定義を新たに生成する。

本節で提案する方式を利用する事で、1つのフィードバック情報から複数のプロパティ制約定義の生成が可能となる。

#### 4. タウンページデータの利用

提案方式をSemantic iタウンページ[7]へ適用する事を検証するため、インターネットタウンページのデータを用いてシステム構築を行った。インターネットタウンページとは、Web上で全国約1,100万件の店舗/会社情報を検索できるサイトの事である。業種は約1,800種あり、それぞれの店舗は必ず一つ以上の業種と結び付けられている。

##### 4.1 Semantic iタウンページとは

Semantic iタウンページとは、利用者が入力したキーワードに対してオントロジを用いて関連情報を表示し、店舗情報を検索するというシステムである。インターネットタウンページのデータを利用している。

具体的な構成は、RDFメタデータを管理・流通するコンテキストビューロ[8][9][10]とデータの処理を行うWebアプリケーションサーバからなる。利用者は、Webアプリケーションサーバにアクセスする。各種オントロジを用いることで、利用者は店舗を探す際に思いついたキーワードから業種を特定し（図5のI）、特定した業種の中から、店舗の提供したメタデータや利用者がフィードバックしたメタデータを用いて更に細かい観点で店舗の絞り込みを行うことができる（図5のII）。利用者からのフィードバック情報を用いる絞り込みは、他の利用者のフィードバック情報を参考にでき、それまでに考えていなかった観点から店舗の絞り込みが可能となる。

利用者がフィードバック情報を入力する場合は、プロパティを記入又は選択し、プロパティ値は自由に設定できる。（図5のIV）

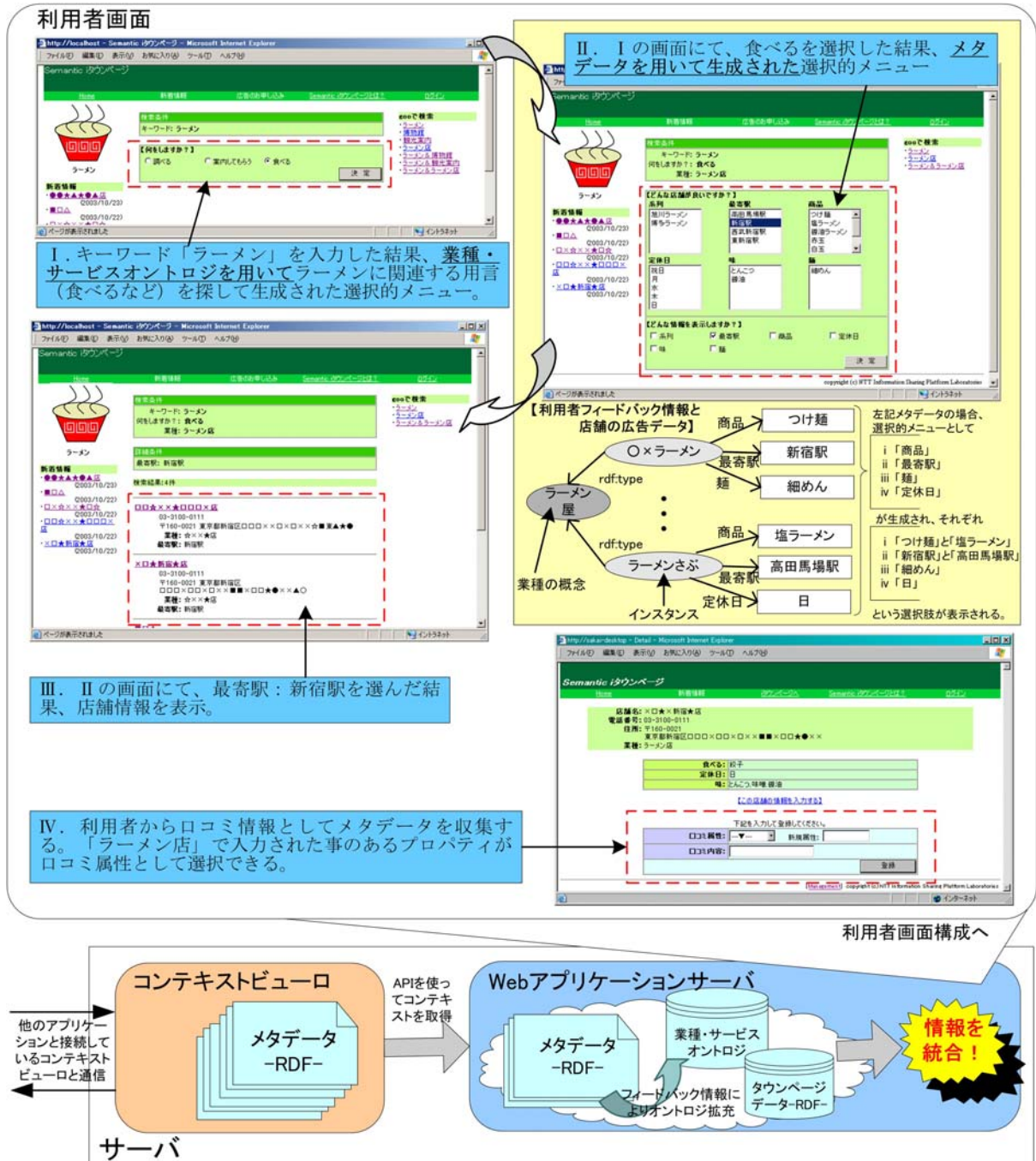


図 5 Semantic i タウンページ

4.2 Semantic i タウンページでの利用データ  
インターネットタウンページの店舗情報の中から、

- ・ 店舗名
- ・ 住所
- ・ 電話番号
- ・ 業種

のデータを RDF 形式に変換して利用した。(図 6)  
また、オントロジとして

- ① リソースのクラス定義  
→業種の上下関係 (図 7)
- ② プロパティ値として利用するクラス定義  
→商品やサービスの上下関係
- ③ プロパティ制約定義  
→domainとして業種、rangeとして商品/サービスを持つプロパティの定義

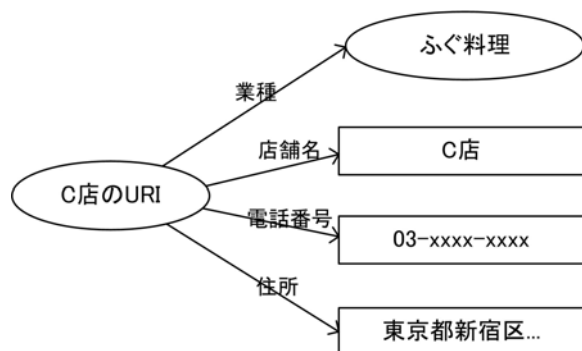


図 6 店舗情報

を用いた。業種の上下関係はインターネットタウンページに用意されている「業種分類」を用いており、オントロジ内では、業種をクラスとして扱っている。サービス・商品のオントロジ及びプロパティの定義は、岩瀬[11]らが作成した「動詞を介した商品と職業の関連」のデータを基に作成した。

## 5. 提案方式の適用と評価

### 5.1 適用

本方式の効果を評価するため、インターネットタウンページのテキストの広告データから、メタデータ抽出及び、利用者からのフィードバック情報を用いたオントロジ拡充を行った。

インターネットタウンページ内の約 1,800 業種から 5 業種を選んだ。

- (ア) うなぎ料理
- (イ) てんぷら料理
- (ウ) ふぐ料理
- (エ) 水道衛生工事保守
- (オ) 制服・作業服

これらの業種のある地域の広告データに対して、以下の処理を行った。

- ① 初期状態のオントロジを用いて広告データからメタデータの抽出及びオントロジ拡充、プール情報の抽出を行った。  
オントロジ拡充に関しては、固有名詞をオントロジ化しないために、抽出されたメタデータにおいて、同一の業種で同じプロパティとプロパティ値毎の組が 2 件以上のものだけ

表 1 対象情報

	対象広告数	プロパティ 制約定義数	抽出メタ データ件数
うなぎ料理	14	6	31
てんぷら料理	17	5	32
ふぐ料理	27	1	78
水道衛生工事保守	70	10	387
制服・作業服	41	5	97

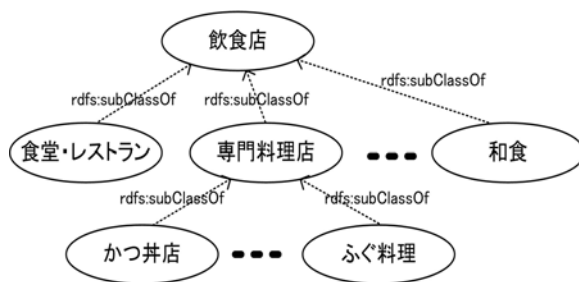


図 7 業種の上下関係

を拡充対象とした。

- ② ①でプールした情報に対して、利用者からのフィードバック情報と想定したリソース(サブジェクト)、プロパティ、プロパティ値の組を入力値として与え、オントロジ拡充を行った。

入力値はプール情報のプロパティ値の候補から 2 または 3 個を選び、プロパティを与えて以下の組(プロパティ, プロパティ値)を作成した。この場合のオントロジ拡充も①と同様に、同一の業種で同じプロパティとプロパティ値毎の組が 2 件以上のものを対象とした。

#### <うなぎ料理>

- (提供する, うなぎ)
- (提供する, 鍋)

#### <てんぷら料理>

- (提供する, 定食)
- (提供する, ビール)

#### <ふぐ料理>

- (提供する, ふぐ)
- (提供する, 白子)
- (提供する, 唐揚げ)

#### <水道衛生工事保守>

- (修理する, パイプ)
- (修理する, 給水管)
- (修理する, 洗面所)

#### <制服・作業服>

- (売る, ユニフォーム)
- (売る, 軍手)
- (売る, 白衣)

表 2 適合率

	抽出メタ データ件数	適合メタ データ件数	適合率
うなぎ料理	31	19.3	0.62
てんぷら料理	32	27	0.84
ふぐ料理	78	59	0.76
水道衛生工事保守	387	239	0.62
制服・作業服	97	83.3	0.86

表 3 オントロジ拡充結果

	拡充件数	拡充オントロジ適合数	適合率
うなぎ料理	6	3.5	0.58
てんぷら料理	5	4.75	0.95
ふぐ料理	11	8.75	0.79
水道衛生工事保守	40	23.25	0.58
制服・作業服	11	9	0.81

## 5.2 評価

対象とする業種の広告数、事前に与えたオントロジのプロパティ制約定義の数、及び 5.1 の①で抽出したメタデータの件数を表 1 に示す。プロパティ制約定義数よりも対象広告数に依存して抽出件数が増加している事が分かる。

抽出したメタデータに関して適合率を評価した結果が表 2 である。4 名の被験者が、メタデータのサブジェクトの業種に対してプロパティとプロパティ値の対応関係が合っているかどうかをチェックした。表 2 での適合率は、4 名の結果を平均したものである。

表 3 はメタデータの適合率を利用して算出したオントロジ拡充の適合率である。固有名詞をオントロジ化しない事により、メタデータの抽出と同程度の結果が得られる事が分かった。

5.1 の①で抽出されたプール情報の数と、5.1 の②の利用者フィードバック情報を利用して生成されたメタデータの数とオントロジ拡充数を表 4 に示す。特に制服・作業服は 3 件のフィードバック情報から 12 件のプロパティ制約定義がオントロジへ追加されている。

## 6. 考察

インターネットタウンページの広告データからプロパティ値候補の抽出方法は、形態素解析で品詞を判定する程度のものである。しかし、抽出データの結果を見ると、この程度のアルゴリズムでも平均 7 割程度の適合率になる。例えば、プロパティ制約定義が 1 つしかない「ふぐ料理」では、「ふぐ料理では、ふぐ料理を提供する」という制約が定義されていた。この定義から、「ふぐ刺しを提供する」や「ふぐちりを提供する」など多数のメタデータが生成された。これは、対象とした文書が広告データで、箇条書きなどが多いことが理由として考えられる。

適合率の低かった「水道衛生工事保守」の抽出結果を見ると、「水道管修理を修理する」の様に「修理」という言葉が重なるなど単語の切り出し方に問題がある例が多数見られた。

メタデータ抽出時のオントロジ拡充データの

表 4 フィードバック情報による拡充結果

	プール情報件数	メタデータ生成件数	拡充オントロジ件数
うなぎ料理	55	13	1
てんぷら料理	74	13	1
ふぐ料理	333	31	7
水道衛生工事保守	2079	102	4
制服・作業服	1016	30	12

適合率は抽出メタデータの適合率に比例する結果が得られている。今後は、自然言語解析技術を上手く取り入れる事を検討していきたい。

利用者フィードバック情報の利用に関しては、プール情報に存在するデータが入力されれば関連したメタデータが生成され、オントロジの拡充も行われた。

## 7. おわりに

インターネットタウンページ上の広告データを利用し、提案方式によるメタデータ抽出実験及び、オントロジ拡充実験を行い、プール情報を扱う事で、1 つのフィードバック情報から複数のプロパティ制約定義をオントロジへ追加できる事が確認できた。

しかし、利用者のフィードバック情報からだけでは新しい概念の上下関係までは解らないという問題がある。また、本方式では、フィードバック情報の質が重要になるため、プール情報の中からプロパティ値の候補を表示したり、該当業種に関連するプロパティ候補を表示したりするなど、ナビゲーション的な要素を入れ、フィードバック情報の信頼性を保つ事が今後の研究課題の一つである。

## 参考文献

- [1] goo (<http://www.goo.ne.jp>)
- [2] インターネットタウンページ (<http://itp.ne.jp/>) NTT 番号情報株式会社
- [3] TAP (<http://tap.stanford.edu/>)
- [4] Graham Klyne, Jeremy J. Carroll, Resource Description Framework (RDF) : Concepts and Abstract Syntax (<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>)
- [5] Dan Brickley, R.V. Guha, RDF Vocabulary Description Language 1.0: RDF Schema (<http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>)
- [6] 松尾豊 福田隼人 石塚満, ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援, 人工

知能学会論文誌 18 巻 4 号 E pp. 203-211, 2003

[7] 境美樹 佐藤宏之 中村昌志, Semantic i タウンページ, INTAP セマンティック Web コンファレンス 2003

[8] 佐藤宏之, 「セマンティック Web」入門 第5回, 日経インターネットソリューション 8月号 p111-118, 日経 BP 社, 2003

[9] 小倉弘敬 村上佐枝子 佐藤宏之 小島富彦 清水昇 細見格, セマンティック Web の応用システム, 情報処理, vol. 43, No. 7, 情報処理学会, 2002.

[10] 佐藤宏之 飯塚京士 境美樹 中村昌志, コンテキストを流通するプラットフォーム技術とアプリケーション, INTAP セマンティック Web コンファレンス 2003

[11] 岩瀬成人 大山実, 自然言語処理技術を用いた職業別電話帳検索の高度化, 電子情報通信学会論文誌 Vol. J74-D-II No. 9 pp. 1255-1263, 1991