

オントロジを利用した情報パーソナライズ、カスタマイズ方法の提案

伊藤浩二

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

ito.koji@lab.ntt.co.jp

あらまし： Semantic Web を用いることにより、コンテンツと概念を正確に関連付け、情報を過不足無く検索する事が期待されている。本稿ではユーザのプリファレンス情報を用い、オントロジの各リンクに設定した強度を更新し、更新された強度を用いて情報を選択する方法を提案する。

1 はじめに

Web 上に存在する情報の増加に伴い、ユーザが自らに適する情報を探し出す事はますます困難となっている。そのため、ユーザの要望やプリファレンス情報、あるいは情報送信元の要望に基づき、コンテンツの重要度を決定し、情報を選択する技術の重要性が増している。従来はこのような要求に対し、パーソナライゼーション、あるいはカスタマイゼーションが用いられてきた。これら従来技術に対し、W3C では Semantic Web の適用例として、趣味嗜好などのプリファレンス情報を用いて情報を選択し、ユーザに提供するインテリジェントエージェントについて述べている[1]。

Semantic Web を用いることにより、情報選択の精度を向上させることが期待されている。本稿ではオントロジを用いて、パーソナライゼーション及び、カスタマイゼーションを拡張する方法を提案する。また、提案方法について、プロトタイプを構築した結果についても報告する。

2 従来技術

2.1 パーソナライゼーション

パーソナライゼーション(図1)は、ユーザのプリファレンス情報から、ユーザに対し適する情報を自動的に選択する方法であり、ルールベース、協調フィルタリング等の手法が用いられる。

ルールベースは、既知の、あるいはデータマイ

ニング等の手法により発見された配送ルールを情報配送時に指定することにより、情報のパーソナライズを行う方法である。

協調フィルタリングは、類似の評価を行っている他者の嗜好情報を基に、ユーザの嗜好を推測する方法であり、ユーザの利用履歴等から自動的にリコメンド情報を作成可能、運用が簡単であるという特徴を持つ。しかし、管理者によるルールの追加変更ができないという課題が Eliseo 等により指摘されている[2]。

協調フィルタリングの仕組み上、本課題の解決は非常に困難であるため、本稿では、管理者によるルールの追加変更が可能なルールベースに着目し、課題解決を目指す。

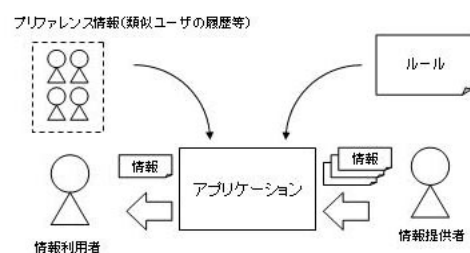


図1 パーソナライゼーション

2.2 カスタマイゼーション

カスタマイゼーション(図2)は、ユーザが入力した情報に基づき、画面レイアウトや情報を選択する方法であり、Opt-in/Opt-out等の手法が用いられる。例えば、柏崎市が運用するマイページ

はカスタマイゼーションを実現した例であり、市民が予め登録した情報に基づき、適する情報が表示される[3]。Opt-in/Opt-out は比較的容易に実現できるため、現在一般的に使われている。

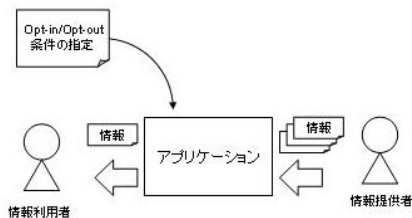


図 2 カスタマイゼーション

2.3 従来技術の課題

以下に従来技術を用いてコンテンツの重要度を決定する際の課題を述べる。

ルールベースでは、ルールは専門家が経験的に有していたもの、あるいは、データマイニング等により導かれたものであり、追加変更は困難である。また、多くのルールを扱う場合は計算量が増加し、リアルタイムな情報の絞込みが困難であり、用いるプリファレンス情報に所在地など、更新頻度が速いものが含まれる場合は、予め計算することも困難である。また、ユーザに関する情報を元にルールを実行する場合は、ユーザに関する情報が必須である。さらに、協調フィルタリングと異なり、ルールを手動で登録する必要がある。

Opt-in/Opt-out では、ユーザの要求は多種多様であり、ユーザが満足する結果を得るためには、非常に多くの Opt-in/Opt-out の設定を行う必要がある。例えば、マイページの例では、「出産」「育児」「就学」などから選択するが、ユーザのニッチな要求に答えるために、選択肢を増やすほど使い勝手が低下する恐れがある。また、ユーザの要求は“要求する”あるいは、“要求しない”という二値的なものではなく“強く要求する”あるいは、“あればあったて良い”というような曖昧な要望、あるいはそれらの組み合わせであることが多いが、このような要求にこたえる事はできない。

以下に従来技術の課題をまとめる。

◇ ルールベースの課題

1. ルールの追加変更が困難
2. リアルタイムな応答が困難
3. ユーザ情報が必須
4. ルールを手動で登録する必要がある

◇ Opt-in/Opt-out の課題

5. Opt-in/Opt-out 指定対象の限定
6. 要求程度の指定が不可能

3 提案手法

Semantic Web において、概念間の関係を記述したものをオントロジと呼び、オントロジ記述言語(OWL)が標準化されている[4]。図 3 にオントロジの例を示す。本稿では、オントロジを用い、コンテンツの重要度を決定する。さらに、これを拡張することにより、Opt-in/Opt-out によるカスタマイゼーション及びルールベースによるパーソナライゼーションを実現する。

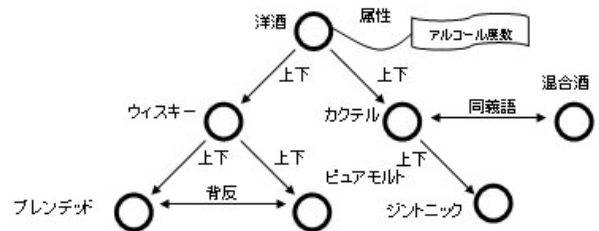


図 3 オントロジ

具体的には、概念(クラス、以下 Cl_n と表す)間、概念-コンテンツ間の関連(リンク)にリンク強度を定め、有向グラフであるオントロジを、重み付き有向グラフ(ネットワーク)に拡張する。各リンクには、プリファレンス情報を用いてリンク強度を更新するルールを関連付け、リンク強度を適宜更新する。一方、ユーザは要求の程度を伴う Opt-in/Opt-out 条件を指定する。指定された条件に対し、オントロジに蓄積されたリンク強度を用い、概念及びコンテンツの重要度を算出する。

以下、オントロジの拡張、利用方法、及び提案手法の効果について説明する。

3.1 リンクの拡張

提案手法では、クラス間及び、クラス-コンテンツ間のリンクに着目し、リンクの強度により、概念、コンテンツの重要度を判断する。以下、リンク及びリンク強度の定義について述べる。

3.1.1 クラス間のシナプス

先ず、図4に示す通り、クラス間のリンクを表すクラス(以下クラス間シナプスと呼び、 $Syn(Cl_n, Cl_m)$ と表す)を定める。このクラス間シナプスにリンク強度を定め、この値によってクラス間の関連の度合いを表す。(以下クラス間リンク強度と呼び、 $Int(Syn(Cl_m, Cl_n))$ と表す)

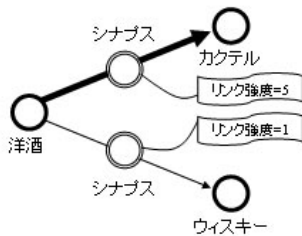


図4 クラス間シナプスの定義

例えば、図4の例では、

$$Int(Syn(Cl_{洋酒}, Cl_{カクテル})) = 5 \quad (1)$$

$$Int(Syn(Cl_{洋酒}, Cl_{ウイスキー})) = 1 \quad (2)$$

であり、カクテルはウイスキーより洋酒からの関連が強い事を表す。

3.1.2 クラス - コンテンツ間シナプス

図5に示す通り、クラスとコンテンツの間の関連を表すクラスを定める。(以下クラス - コンテンツ間シナプスと呼び、 $Syn(Cl_m, Co_n)$ と表す) このクラス間シナプスにリンク強度を定め、この値によってクラス、コンテンツ間の関連の度合いを表す。(クラス-コンテンツ間リンク強度と呼び、 $Int(Syn(Cl_m, Co_n))$ と表す) 例えば、図5の例では、 $Int(Syn(Cl_{カクテル}, Co_1)) = 5$ (3)

$$Int(Syn(Cl_{カクテル}, Co_2)) = 1 \quad (4)$$

であり、コンテンツ1はコンテンツ2よりカクテルからの関連が強い事を表す。

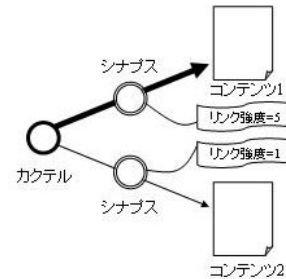


図5 クラス - コンテンツ間シナプスの定義

3.2 リンク強度の更新

提案手法では、各シナプスに対し、ルールを定義し、リンク強度を更新する。以下、単純なルールの組み合わせにより、リンク強度を更新する方法について述べる。

3.2.1 ルールクラスの定義

図7に示す通り、シナプスに対し、リンク強度を更新するルールを表すクラス(以下ルールクラスと呼び R_n と表す)を定める。

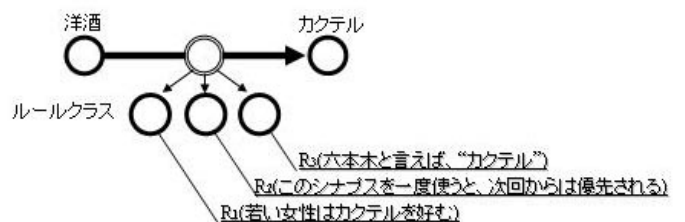


図6 ルールクラスの定義

例えば、図6の例では

R_1 : ユーザの性別が“女性”である

R_2 : ユーザがこのシナプスを利用する

R_3 : ユーザ所在地が“六本木”近辺である

以上のいずれかの条件を充たす時、洋酒とカクテルのリンク強度が更新される。

個々のルールクラスには適用タイミングが定

められる。例えば、性別のように更新頻度が遅いものは、設定時に計算され、所在地のように更新頻度が速いものはアクセス時に計算される。

3.2.2 条件クラスと実行クラス

ルールクラスは、図7に示す通り、再利用可能な程度に細分化された条件クラス及び実行クラスから構成され、それぞれ実際に条件の判断及びリンク強度の更新を行うオブジェクト指向言語のクラスを関連付ける。

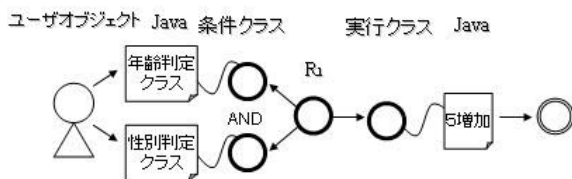


図7 条件クラスと実行クラス

3.3 リンク強度の融合

リンク強度は図8に示す通り、ユーザの指定した割合(リンク融合率)に応じ、ユーザ間で共通に用いるオントロジ(ベースオントロジ)のリンク強度とユーザに依存するオントロジ(個人オントロジ)のリンク強度を融合し用いる。

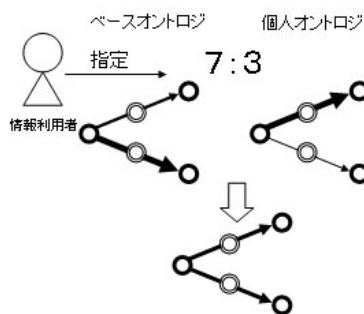


図8 オントロジの融合

3.4 Opt-in/Opt-out クラス

3.2、3.3節で求めたリンク強度に従い、コンテンツをユーザごとに選択し提供する場合、ルールベースのパーソナライゼーションの一種と考え

ることが可能である。3.4節ではこれに加え、ユーザのニッチ、曖昧な要望を指定可能な、カスタマイゼーションについて説明する。

ユーザは、図9に示す通り、ユーザはオントロジに含まれる概念から、関連する情報の取得を望む概念、及び望まない概念(Opt-in/Opt-out クラス、以下 cl_n^{opt} と表す)を発見し、指定する。この際、ユーザからのリンク強度が高い概念を優先的に表示する。例えば、図9はユーザが宿泊施設の種別を Opt-in/Opt-out クラスに指定する例であるが、旅館(リンク強度=10)は貸し別荘(リンク強度=3)と比較し、優先的に Opt-in/Opt-out クラス候補として表示される。これにより、オントロジに含まれる多くの概念から Opt-in/Opt-out クラスを効率的に発見することが可能となる。

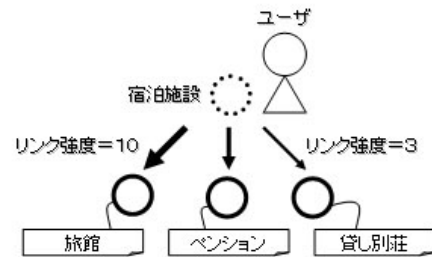


図9 Opt-in/Opt-out クラスの指定

ユーザは、図10に示す通り、Opt-in/Opt-out クラスを指定する際、Opt-in/Opt-out の程度 (Opt-in/Opt-out 率、以下 $ratio(cl_n^{opt})$ と表す) を指定可能である。図10では宿泊施設を検索する場合の Opt-in/Opt-out クラス指定例であるが、

- ◇ “スキー” に対する宿を強く望む
- ◇ できれば北海道が良い
- ◇ ペットと泊まれる宿は避けたい

という希望をユーザが表現したことになる。

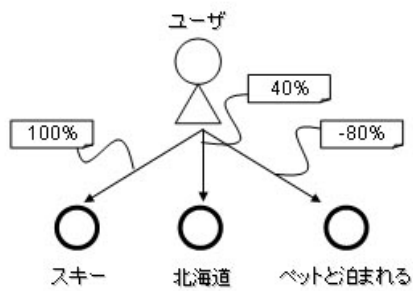


図 10 Opt-in/Opt-out 率の指定

4 提案手法の効果

3.1 節から 3.4 節で述べた提案手法の効果を以下にまとめる。

4.1 管理者によるルールの追加変更

3.2.1 項に示した通り、本提案で扱うルールは、個々の概念間のリンク強度あるいは、概念とコンテンツ間のリンク強度を更新する比較的単純なものであり、さらに、3.2.2 項で述べた通り、条件、実行、Java クラスはそれぞれ再利用が可能な程度に細分化されているため、ルールの追加変更が容易である。従って、追加したルールの効果を推定することが困難であるものの、ルールベースにルールを追加する場合と比較し、ルールクラスを簡単に追加変更可能となる。

4.2 リクエストに対する即時応答性

3.2.1 項に示した通り、更新頻度が遅いプリファレンス情報については予め計算を行い、計算結果をリンク強度として蓄積することにより、リクエスト発生時の計算量を抑制する。従って、更新頻度が遅いプリファレンス情報の割合が大きい場合、リアルタイムな応答が可能となる。

4.3 ユーザ情報が少ない場合の検索精度

3.3 節に示した通り、コンテンツの検索にあたり、ユーザ間で共通に用いるオントロジ(ベースオントロジ)のリンク強度とユーザに依存するオントロジ(個人オントロジ)のリンク強度を融合し使い

ることが可能である。これにより、プリファレンス情報が少ない場合、個人に対する適応率は低いものの、検索結果を得ることが可能となる。

4.4 Opt-in/Opt-out 指定対象の拡張

3.4 節で述べたとおり、オントロジに含まれる任意の概念を Opt-in/Opt-out クラスとして指定可能であり、詳細な条件の指定が可能となる。

4.5 要求程度の指定

3.4 節で述べたとおり、Opt-in/Opt-out クラスには、Opt-in/Opt-out 率の指定が可能であり、詳細な条件の指定が可能となる。

5 情報検索への応用

図 11 にシナプス、リンク強度、Opt-in/Opt-out 条件を用い情報検索を行う例を示す。

5.1 Opt-in/Opt-out クラスの特定

ユーザが指定した Opt-in/Opt-out クラス及び、同値、 subclasses を Opt-in/Opt-out クラスとする。例えば、図 11 においてユーザがクラス A, B, C を Opt-in/Opt-out クラスに指定した場合、クラス D ~ クラス I も Opt-in/Opt-out クラスに含む。

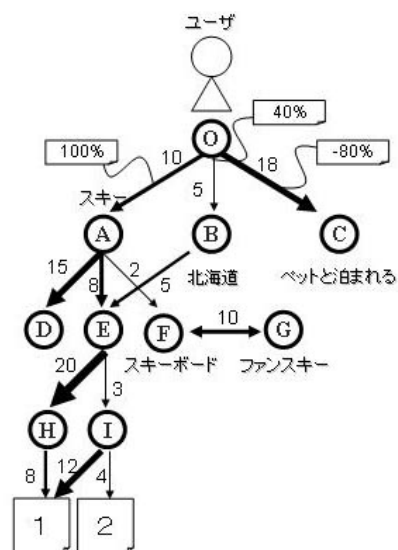


図 11 情報検索処理

5.2 木構造への変換

複数の親を持つクラスは、親クラスごとに別のクラスとして扱い、木構造に変換する。例えば、図 10 において、クラス E はクラス E (クラス A 経由) とクラス E (クラス B 経由) に展開され、クラス E が子ノード、同値ノードを持つ場合も、別のクラスとして扱う。例えば、クラス H は、クラス H (クラス A 経由) とクラス H (クラス B 経由) として扱う。

5.3 リンク強度の更新

4.4 で示した方法に従いリンク強度を更新する。この際、子ノード、同値ノードに対してもルールを適用する。例えば、

◇ 冬にはスキーが好まれる

というルールが図 11 に示すクラス O、クラス A 間のシナプスに定められている場合、このルールは、

◇ 冬にはスキーボードが好まれる

◇ 冬にはファンスキーが好まれる

というルールとして、子クラス、同値クラスへのシナプスにも適用される。

5.4 Opt-in 率/Opt-out 率の決定

ユーザが指定した Opt-in 率/Opt-out 率を子ノード、同値ノードに対して継承する。例えば、クラス A の Opt-in 率は 100%、クラス B の Opt-in 率 40% である場合、クラス K の Opt-in 率(クラス A 経由)は 100% であり、クラス K の Opt-in 率(クラス B 経由)は 40% である。

5.5 クラス重要度の決定

当該クラスに至ったリンクのリンク強度に Opt-in 率/Opt-out 率を乗じたものをクラス重要

度(以下 Inp_{Cl_n} と表す)と定める。

$$Inp_{Cl_n} = Int(Syn(Cl_m, Cl_n)) \cdot ratio(Cl_m^{opt}) \quad (5)$$

例えば、図 11 においてクラス H(クラス A 経由)の重要度は、

$$\begin{aligned} Inp_{Cl_H} &= Int(Syn(Cl_{E(A)}, Cl_H)) \times ratio(Cl_{E(A)}^{opt}) \\ &= 20 \times 1.0 \\ &= 20 \end{aligned} \quad (6)$$

と求める。

5.6 コンテンツ重要度の決定

当該コンテンツに対しリンクを持つすべてのクラスについて、クラス重要度、及び当該コンテンツへのリンク強度と Opt-in/Opt-out 率の積を加え、コンテンツの重要度(以下 Inp_{Co_n} と表す)と定める。

$$Inp_{Co_n} = \sum_m (Inp_{Cl_m} + Int(Syn(Cl_m, Co_n)) \times ratio(Cl_m^{opt})) \quad (7)$$

例えば、図 11 において、コンテンツ 1 の重要度は以下のように求める。

$$\begin{aligned} Inp_{Co_1} &= Inp_{Cl_{H(A)}} + Inp_{Cl_{H(B)}} + Inp_{Cl_{I(A)}} + Inp_{Cl_{I(B)}} \\ &+ Int(Syn(Cl_{H(A)}, Co_1)) \times 1.0 + Int(Syn(Cl_{H(B)}, Co_1)) \times 0.4 \\ &+ Int(Syn(Cl_{I(A)}, Co_1)) \times 1.0 + Int(Syn(Cl_{I(B)}, Co_1)) \times 0.4 \\ &= 20 + 8 + 3 + 1.2 + 8 + 3.2 + 12 + 4.8 = 60.2 \end{aligned} \quad (8)$$

以上の通り算出したコンテンツ重要度を用い、ユーザに提供する情報を選択する。

6 提案手法の実装例

本稿で提案した情報パーソナライズ方法の実現可能性を実証するため、提案した手法を宿泊施設リコメンドシステムに適用した。

プロトタイプ構築にあたっては、オントロジ管理エンジンとして、Stanford 大学が開発した Protégé[5]を用いた。

6.1 Opt-in/Opt-out クラスの指定

ユーザは、Opt-in/Opt-out クラス指定画面より、Opt-in/Opt-out クラスを指定する。画面上に表示される選択肢はユーザごとに管理されたオントロジに保存されたリンク強度を用いてソートされる。例えば、ユーザがある選択肢をクリックした場合、選択された概念に対しユーザが興味を持ったという事実を元にリンク強度を更新し、再度ソートされる。



図 12 Opt-in/Opt-out クラス指定画面

6.2 検索結果の表示

ユーザが指定した Opt-in/Opt-out クラスを 5.1 に示す通り、オントロジを用いて拡張し、最終的な Opt-in/Opt-out クラスとする。図 13 では、ユーザが“ホテル”、“宮崎県”、“夫婦で”、“10000円～15000円”という4つの概念を指定した場合であるが、例えば、ホテルの子概念として“シティホテル”が存在する場合はこの概念も Opt-in/Opt-out クラスとなる。

ユーザはその他に、3.3 及び 3.4 で述べたリンク融合率及び Opt-in/Opt-out 率を指定する。図 13 は、リンク融合率を 1:1、Opt-in/Opt-out 率をすべて 0.5 とした場合の結果である。検索結果

はコンテンツ重要度に応じてソートされ表示される。



図 13 検索結果画面

Opt-in/Opt-out クラス、リンク融合率、Opt-in/Opt-out 率のいずれかが変更された場合は直ちに再検索され、検索結果が更新される。サムネイルをクリックすると各コンテンツのコンテンツ重要度、そのコンテンツに対してリンクを持つ概念をリンク強度に応じてソートしたものが詳細情報として表示される。

6.3 Opt-in/Opt-out クラスの追加

図 14 に示す通り、検索されたコンテンツに対してリンクを持つ概念近傍のオントロジを表示することができる。この際、リンク強度が強いシナプスは太い線で、弱いシナプスは細い線で描画される。ここで発見した概念を Opt-in/Opt-out クラスに加えることにより、ユーザは徐々に検索条件を追加編集し、詳細化することができる。

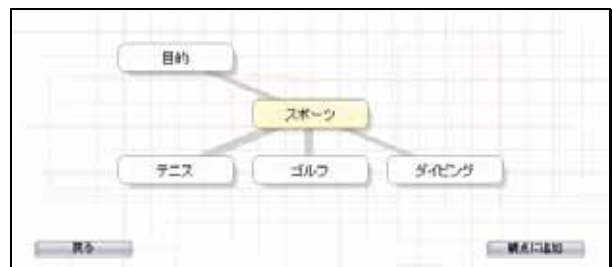


図 14 コンテンツに関連するオントロジの表示

6.4 プロトタイプの評価

プロトタイプにより、オントロジの各リンクに設定された強度を更新し、更新された強度を用いて情報を選択する方法の実現可能性が確認された。しかし、プロトタイプで利用したオントロジはクラス数が 50 程度と小さく、用いたルールの種類も少ない。今後、提案手法が効果的な適用分野を見極めた上で、実運用に必要なオントロジ規模、システム負荷、情報管理方法、個人情報収集、管理方法等について検討を行う必要がある。

7 まとめ

本稿ではユーザが設定したプリファレンス情報を用い、オントロジの各リンクに設定された強度を更新し、更新された強度を用いて情報を選択する方法を提案した。さらに、プロトタイピングを行い、プリファレンス情報をリンクに蓄積し、ユーザの曖昧な Opt-in/Opt-out 要求に対し、逐次リコメンデーション結果を表示する仕組みの実現可能性を確認した。

本手法は、Opt-in/Opt-out 項目の追加、及び Opt-in/Opt-out 率の指定により、詳細かつ柔軟なカスタマイゼーションが可能である。また、リンクごとに定められたリンク強度更新ルールを用い、プリファレンス情報を予めリンク強度に変換し格納し、コンテンツの優先度を計算することにより、ルールの追加変更が容易となり、ユーザに関する多くのプリファレンス情報を用いたパーソナライゼーションも可能となる。

一方、リンク強度を融合する場合は、両者が同じ構造を持つオントロジであること、ルールを手動で発見、登録する必要があること、等の課題もあり、今後解決する必要がある。

本稿で提案した方式には、定量的な評価尺度が存在しない。従って、上記課題を解決した後、ケーススタディを行い、コンテンツ管理の容易さ、スケーラビリティ等、様々な観点より提案方式の有効性を検証する必要があると考えている。

参考文献

- [1] Jeff Heflin. "OWL Web Ontology Language (OWL) Use Cases and Requirements". W3C Recommendation. W3C, 2004-02-10. (online), available from <http://www.w3.org/TR/webont-req/>, (accessed 2005-03-02).
- [2] Eliseo Reategui, John A. Campbell, and Roberto Torres. "Using Item Descriptors in Recommender Systems". AAI Workshop on Semantic Web Personalization 2004. AAI. (online), available from <http://maya.cs.depaul.edu/~mobasher/swp04/accepted.html>, (accessed 2005-03-02).
- [3] 柏崎市. "マイページ". 柏崎市ホームページ, 2003, (online), available from <http://www.city.kashiwazaki.niigata.jp/webapps/mypage/aboutmypage.jsp>, (accessed 2005-03-02).
- [4] Peter F. Patel-Schneider, Pat Hayes, and Ian Horrocks. "OWL Web Ontology Language Semantics and Abstract Syntax". W3C Recommendation. W3C, 2004-02-10. (online), available from <http://www.w3.org/TR/owl-semantics/>, (accessed 2005-03-02).
- [5] J. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubézy, H. Eriksson, N. F. Noy, and S. W. Tu. "The Evolution of Protégé: An Environment for Knowledge-Based Systems Development". SMI, 2003. (online), available from http://www.smi.stanford.edu/pubs/SMI_Reports/SMI-2002-0943.pdf, (accessed 2005-03-02).