

領域オントロジーと個人嗜好情報に基づくモバイル型レコメンデーションシステムの構成

安随 晋太郎[†] 濱崎 雅弘^{‡ ††} 大向 一輝^{‡ ††} 武田 英明^{†† ‡} 山口 高平[†]

[†]慶應義塾大学 〒223-8522 神奈川県横浜市港北区日吉 3-14-1

[‡]総合研究大学院大学 〒101-8430 東京都千代田区一ツ橋 2-1-2

^{††}国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†] {anzui, yamaguchi}@ae.keio.ac.jp, [‡] {hamasaki, i2k}@grad.nii.ac.jp, ^{††} takeda@nii.ac.jp

あらまし 本稿では、領域オントロジーと個人嗜好情報を利用することにより飲食店を対象とした携帯型情報端末用レコメンデーションシステムを提案する。領域オントロジーはディレクトリサービスを対象とし、個人嗜好情報は携帯端末からの位置情報やカテゴリ化されたブックマーク等を用いる。提案システムでは、エンドユーザが屋外で負担なく周辺情報を得られるような検索システムを目指し、ユーザの個人嗜好情報を領域オントロジーによって拡張することにより、見知らぬ土地においてユーザの嗜好に基づいたレコメンデーションを行う。

キーワード オントロジー, セマンティックウェブ, ユーザ適応, モバイル, レコメンデーション

A Mobile Recommendation System with a Domain Ontology and User Preference

Shintaro ANZUI[†] Masahiro HAMASAKI^{‡ ††} Ikki OHMUKAI^{‡ ††} Hideaki TAKEDA^{†† ‡}
and Takahira YAMAGUCHI[†]

[†] Keio University Faculty of Science and Technology 3-14-1 Hiyoshi Kohoku-ku, Yokohama-shi Kanagawa-ken, 223-8522 Japan

[‡] The Graduate University for Advanced Studies 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

^{††} National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: [†] {anzui, yamaguchi}@ae.keio.ac.jp, [‡] {hamasaki, i2k}@grad.nii.ac.jp, ^{††} {takeda}@nii.ac.jp

Abstract Here in this paper, we discuss how to build up a recommendation system for PDA, using a domain ontology and individual preference. The domain ontology comes from a commercial directory service. Individual preference comes from positioning data and categorized bookmarks. Our prototype system should support a user in finding the restaurants that the user will like better. In typical cases, when a user is strange at some area, our prototype system gives the recommendation on restaurants to the user, extending user's individual preference with a domain ontology.

Keyword Ontologies, Semantic Web, user adaptation, mobile, recommendation

1. はじめに

近年、インターネットの普及により多くの情報がウェブ上に溢れ、ユーザが得る情報量は著しく増加した。インターネット上の情報には実世界の事物やサービスと関係の深いものも多く、そうした情報はポータルサイトを通して屋外からでもアクセスすることができる。例として、“楽天-旅の窓口*1”や“ぐるなび*2”などのポータルサイトが挙げられる。こうしたポータルサイトを利用すれば自分の位置情報を利用して最寄りの宿泊先やレストランなどの周辺情報を検索することができる。また、これらの実世界の情報とその地理的な結び

つけを POI(Point Of Interest, 関心地点)として扱うことにより、ナビゲーションなどの近接情報として利用したり、地図上へ表示したりすることも可能である。このように携帯型情報端末などを用いて実世界に関する情報を容易に獲得できる環境が整いつつある。

一方、こうした周辺情報検索には、携帯型情報端末の表示に適したディレクトリ型のサービスが用いられていることが多く、ディレクトリサービスで用いられているディレクトリ構造は一般的に固定的で、巨大なものが多い。

しかし、ユーザごとの興味・関心に関するディレクトリ構造は多様であり、ディレクトリサービスが提供している汎用的かつ固定的な構造では全ての要求に対

*1 <http://www.mytrip.net/guide/keitai.html>

*2 <http://www.gnavi.co.jp>

処できない。例えば、ディレクトリサービスではひとつのクラスに分類されているインスタンス群もユーザの嗜好によっては全く異なる分類をされてしまう(図1)。また、ディレクトリサービスの提供するディレクトリ構造は巨大であることから、期待する結果を得るまでのプロセスが長くなってしまふ。そのためディレクトリ探索には大きなコストが必要になってしまう。

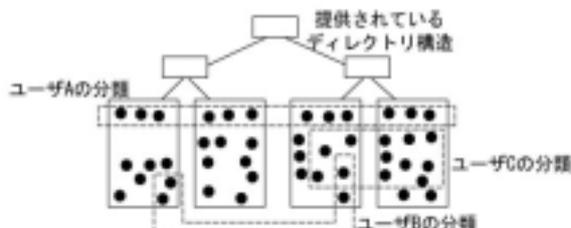


図 1: ユーザごとの分類例

こうした問題点のもと、操作性の低い携帯型情報端末において検索過程や検索結果をユーザの要求や関心に合わせて適切に提供することは、検索プロセスの短縮とユーザの負担軽減につながるものと考えられる。

以上の背景より、本研究では、飲食店情報を対象として領域オントロジーと個人嗜好情報を適用させた携帯型情報端末用レコメンデーションシステムを提案する。提案システムでは、エンドユーザが屋外で負担なく周辺情報を得られるような検索システムを目指し、ユーザの個人嗜好情報を領域オントロジーによって拡張することにより、見知らぬ土地においてユーザの嗜好に基づいたレコメンデーションを行う(図2)。ユーザが作成する既存の世界の情報を活用した嗜好情報はユーザ個人の嗜好体系を反映していると考えられる。しかし、個別に作成されるユーザの嗜好情報は小規模であるため、プロフィールとして情報が不十分である。そこで、多くの情報を持つ領域オントロジーを併用することにより、コストをかけずにユーザの個人嗜好情報を拡張させ、新しい地域でもユーザの嗜好に基づいたレコメンデーションを行えるようにする。

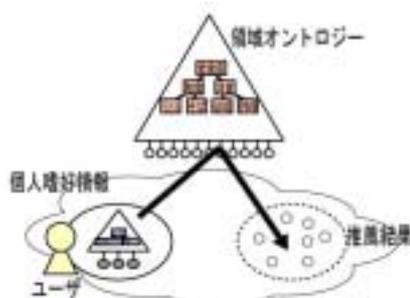


図 2: レコメンデーション概要図

以下、2章で関連研究について、3章では領域オントロジーと個人嗜好情報の適用、4章では実装システムについて述べる。そして、5章で本研究のまとめと

今後の課題について述べる。

2. 関連研究

2.1. セマンティックウェブ

セマンティックウェブ[1]はウェブの拡張であり、情報に明確な意味を与えることで情報の選択や収集を容易に行えるようにして、人とコンピュータの協調実現を目指すものである。この次世代ウェブの実現に向けた議論は World Wide Web Consortium (W3C) を中心に進められており、Resource Description Framework (RDF) [2] や Web Ontology Language (OWL) [3] などをもとに様々な環境でデータの共有・再利用を実現するための標準的なフレームワークづくりが進められている。セマンティックウェブにおいてオントロジーは、概念を詳細に記した知識構造をさす用語として使われており、Gruber の定義[4]では、複数のシステムが動作するために必要な概念や関係を記述するための語彙をフォーマルに記述し、共有・再利用できるようにした知識をオントロジーと呼んでいる。こうしたオントロジーの多くはインターネット上に公開されており、DAML Ontology Library[5] 経由で参照したり、swoogle[6]で検索したりすることが可能となっている。また、オントロジーの統合や参照に関する研究も行われており、伊藤[7]はオントロジーの統合利用などの研究を通してオントロジーの知識共有/再利用の問題への可能性や有効性について述べている。

このようにセマンティックウェブを実現させる環境は徐々に整いつつあり、今後は研究分野のみならず Yahoo!などの商用ディレクトリサービスの分野にも活用されていくものと考えられる。本研究では商用サービスとして提供される知識体系を領域オントロジーとして利用している。

2.2. ユーザ適応

個人のもっている情報はコンピュータが理解できる形として提供されておらず、ユーザはコンピュータを利用する際の経験的にそれらを利用している。たとえば、ウェブ上で検索エンジンを利用する際に、ユーザは自分の必要な情報を得るために適切なキーワードなどの条件を考える。そして、その条件によって得られた検索結果することで取捨選別して自分に必要な情報のみを取得する。こうした個人の知識や嗜好を表すものはユーザプロフィールとされ、さまざまな研究が行われてきた。ユーザ適応をおこなうためには、適応対象となるユーザに関する嗜好や知識体系などの情報(プロフィール)が必要となる。プロフィールを獲得する方法としては、ユーザが直接プロフィールを作成したり、アンケートに入力したりするなどの明示的な手法と、操作履歴などに基づく非明示的な手法が用いられている。

明示的な手法としては以下のようなものがあげられる。佐保田ら[8]や濱崎ら[9]は、ブックマークをユーザの嗜好情報として捉え、推薦をおこなう研究をおこなっている。これらの研究ではウェブページのブックマークを個人嗜好の分類体系として捉え、利用者間の関係や分類体系をもとにして新たなウェブページの推薦をおこなうというものである。

非明示的なものとしては以下のものが挙げられる。官上ら[10]は、ユーザ適応を実現するためのフレームワーク A3 を提案している。A3 では、ユーザが各種のウェブシステムから得る知識をオントロジーにより、体系化して、それをプロファイルとして用いている。また、オントロジーをユーザとシステムのインタラクションから動的に構築し共有することで、システムごとの適応精度差異を軽減させている。大島ら[11]は、個人が日々取り扱う文章などから個人オントロジーを作成して、Web 情報検索の支援をおこなう研究をおこなっている。個人オントロジーで表現された知識を基にして、ウェブ検索時に、すでに持っている情報を排除したり、現在関心の高い優先順位を上げたりするなどにより、ユーザ支援を行っている。

プロファイルを用いず、ユーザの検索意図や興味に沿ったカテゴリ構造を提供することで、ユーザの意図する検索を支援する研究も行われている。仲川ら[12]は、固定的な階層構造では階層構造の見通しが悪いことや、ユーザの想定するカテゴリ構造が検索ごとに異なることなどに着目して、検索のたびにその時々のユーザの想定に近い小規模なカテゴリ構造を構築する方法を提案している。

本研究では、ユーザが既存の世界の情報をもとに作成した小規模の個人嗜好情報を領域オントロジーにより拡張させることで、見知らぬ土地でもユーザの嗜好にしたがってレコメンデーションを行うことを目的としている。

3. 領域オントロジーへの個人嗜好情報の併用

提案システムでは領域オントロジーに個人嗜好情報を併用させることでユーザの嗜好体系を拡張させて検索結果を取得する。具体的には、領域オントロジーが持つ多くのインスタンス群を、ユーザが作成した個人嗜好情に基づいて再分類することでレコメンデーションを行う。また、領域オントロジーによるインスタンス群の限定とディレクトリの推論を行うことで携帯型情報端末のような比較的入出力に制限ある環境でも検索プロセスの軽減が行えると考える。

3.1. 領域オントロジー (Domain Ontology)

領域オントロジーはある特定の領域に関する様々なクラスやインスタンス群を提供するものである。飲食店情報を例にとると、クラスは「和食」「洋食」など

といった分類、インスタンスは各クラスに分類されている個別の飲食店情報を指す(図 3)。

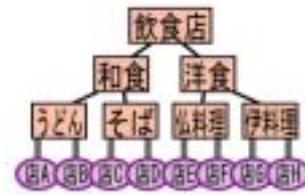


図 3: 飲食店の領域オントロジー例

また、ディレクトリサービスを領域オントロジーとして構築することにより検索時にディレクトリの推論が可能となる。本研究では個人嗜好が顕著に反映され、かつ個人ごとに嗜好の概念が大きく異なる飲食店を領域オントロジーの対象とした。領域オントロジーと個人嗜好情報の適用を実現するためには、飲食店オントロジーを用いる必要があるが、実験に適した飲食店オントロジーが現時点ではまだ提供されていないため、今回は Yahoo! グルメ[13]が提供しているディレクトリ型データベースを対象として RDF と OWL を用いて簡易的に飲食店オントロジーを構築した。これ以降、飲食店オントロジーが提供するクラスを飲食ジャンル、インスタンスを飲食店情報と呼ぶ。

3.2. 個人嗜好情報

本研究では、個人の嗜好体系を表現している個人嗜好情報を利用し、ユーザの嗜好体系に従ったレコメンデーションに利用している。一般にユーザは獲得した情報を整理分類して知識体系を構築していくと考えられるために、本研究ではユーザによって行われる分類自体がユーザの嗜好体系を表すものと考えている。個人による分類は主観的な各自の嗜好に基づいて作成、分類される。例として飲食店情報を例にするとユーザの分類としては「雰囲気がいい」「楽しい」などの分類が考えられる(図 4)。

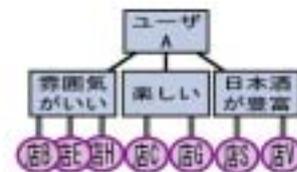


図 4: 飲食店の個人嗜好情報の例

このように作成された、飲食店情報を分類するためのカテゴリ(ユーザカテゴリ)は、検索結果を個人の嗜好に基づいて再分類するために利用している。また個人嗜好情報は追加されていくことで、ユーザカテゴリが増加していき、ユーザの嗜好情報の特徴が明確になっていく。

3.3. 領域オントロジーと個人嗜好情報の適用方法

提案システムでは領域オントロジーと個人嗜好情報の適用を2段階の処理で行う(図5)。

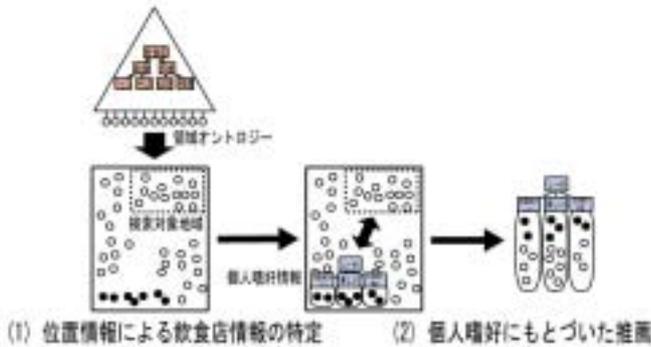


図5: 領域オントロジーと個人嗜好情報の適用

(1) 領域オントロジーの飲食店情報群を利用して位置情報をもとにその地域の飲食店を特定する。

(2) 位置をもとに領域オントロジーから特定された飲食店情報と嗜好情報との類似性を求め、ユーザの嗜好をもとにした分類を行う。

(1)の段階でユーザの位置情報を用いて領域オントロジーからその地域の飲食店を特定する。(2)ユーザが個人嗜好情報として持っている飲食店情報とそのユーザカテゴリをもとに領域オントロジーの飲食店情報との類似性を求め、ユーザに嗜好情報に基づいて分類を行う。

3.4. ユーザカテゴリへの再分類手法

飲食店情報の類似度判定に利用する飲食店コメントコーパスをあらかじめ構築しておく(図6)。



図6: 飲食店コメントコーパスの構築

まず、各飲食店に関するコメントをWeb上から収集する。収集されたコメントは茶筌[14]を利用して形態素解析して品詞を抽出する。さらに専門用語抽出モジュール[15]を利用して専門用語の抽出を行う。抽出した単語列をもとにコメントの特徴ベクトルを作成し、それをもとに飲食店コメントコーパスを構築する。文書の特徴ベクトルはベクトル空間モデル[16]に基づいて単語の頻度を用いて作成した。

ユーザの嗜好情報と検索結果の類似性を比較するため、ユーザカテゴリごとに特徴ベクトルが計算され

る。特徴ベクトルの算出にはカテゴリに分類されている飲食店情報を用いる。ユーザカテゴリの特徴ベクトルはユーザカテゴリ内の全飲食店情報のベクトル和を用いて特徴ベクトルとしている。

3.5. 類似度計算

検索結果の飲食店情報をユーザカテゴリに再分類するための類似度計算は以下のように計算される。

特徴ベクトルをもとに、TFIDF法を拡張したSMARTシステム[17]による類似度算出方法を用いて計算を行う。SMARTにおいてユーザカテゴリ c と飲食店情報 d との類似度を $SMART(c, d)$ とおいたとき、ユーザカテゴリ c と飲食店情報 d との類似度は、 $sim(c, d) = SMART(c, d)$ である。なお、SMARTのアルゴリズムによる類似度の計算にはGETA[18]を用いた。

4. 実装システム

試作システムは飲食店検索部とユーザ適応部、個人嗜好情報編集部で構成されている(図7)。それぞれCGI+Perl5で作成した後、サーバに配置して屋外から携帯型情報端末を用いてアクセスする。利用データは、位置データベースとしてオンライン上から取得した首都圏の路線情報、飲食店データベースとしてYahoo!グルメの首都圏の飲食店情報を用いた。

4.1. システム処理の流れ

- (1) ユーザは携帯情報端末から位置情報を送り最寄り駅の情報を受け取る。
- (2) 最寄り駅に飲食店がない場合、代替検索が行われ、同じ路線の駅が代替駅としてユーザに提示される。
- (3) 飲食店オントロジーの飲食店ジャンルを追加条件として用いたい場合、ユーザは希望する飲食店ジャンルを指定する。
- (4) (3)で飲食ジャンルを選択した際に、最寄り駅に該当する飲食店がない場合、領域オントロジーの分類体系にもとづいて代替検索が行われ新しい飲食店ジャンルがユーザに提示される。
- (5) (1)~(4)で選択された条件から周辺検索が行われ該当する飲食店が結果として得られる。
- (6) 条件に該当する飲食店に対してユーザの嗜好情報が適用され、該当飲食店とユーザカテゴリの類似度が計算される。
- (7) 検索結果に対して類似度が計算された後、ユーザカテゴリとの類似度が高いものに分類される。
- (8) ユーザは検索結果のURIを個人嗜好情報編集部に送ることで個人嗜好情報を追加することもできる。

4.2. 飲食店検索部

飲食店検索部は、位置情報と飲食店ジャンル情報から周辺検索を行う。位置情報による最寄り駅の検索は位置検索部、領域オントロジーによる飲食店ジャンルの選択は飲食店ジャンル選定部によって行われる。

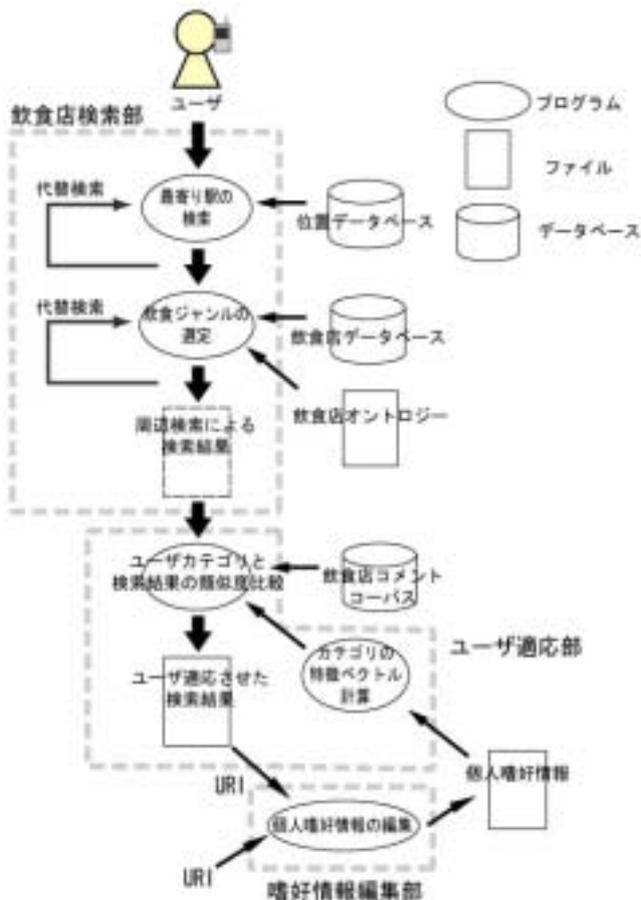


図 7: 試作システムの構成

4.2.1. 位置検索部

位置検索部ではユーザの現在地を基点として最寄り駅を検索する。ユーザは GPS 付携帯型情報端末から飲食店検索部に位置情報を送信する。位置情報は GPS から得た緯度・経度情報を利用した。送信された緯度・経度による位置情報は位置検索部に送られ、駅データベースから最寄り駅の情報が見られる。駅データベースは MySQL4.0 により構築しており、駅名、駅所在地の緯度・経度、鉄道会社、路線、隣接する駅などの情報が格納されている。今回は試作システムであるため駅データベースに登録されている駅情報は関東地方にある一万三千程度の駅を対象とした。最寄り駅の情報が見られた後は、見られた駅情報をもとに飲食店データベースから最寄り駅付近に登録されている飲食店数を調べる。最寄り駅付近に飲食店が存在しない場合、位置検索部はユーザにその旨を伝え隣接する駅を代替駅として提示してユーザから選択された駅を再度位置検索部に送る。代替駅の検索は、駅データベースから最寄り駅の路線を特定し、その路線上で隣接する駅を代替駅としてユーザに提示することで行っている。

4.2.2. 飲食店ジャンル選定部

最寄り駅、もしくは代替駅によって検索対象となる地域が確定した後、ユーザが領域オントロジーの飲食ジャンルによる対象限定を行うかを選択する。もし、ユーザが領域オントロジーの分類も同時に利用する場合は、領域オントロジーのもつ飲食店ジャンル一覧が表示される。表示される分類は RDF+OWL で記述された Yahoo!グルメの領域オントロジーであり、Yahoo!グルメの提供しているディレクトリ構造を提示する。ユーザは表示された飲食店ジャンルの中から検索したいジャンルを選択する。選択された飲食店ジャンルと位置情報の条件をもとに飲食店データベースから条件に該当する飲食店情報を受け取る。飲食店データベースは駅データベース同様 MySQL4.0 により構築されている。データとして飲食店名、住所、電話番号、URL などが格納されている。今回は関東地方にある Yahoo!グルメの飲食店、一万四千店程度を対象とした。

4.3. ユーザ適応部

ユーザ適応部は飲食店検索部によって得られた検索結果に対して個人嗜好情報を適用させて、ユーザカテゴリとの類似性を比較する。

飲食ジャンル選定部から送られてきた検索結果はユーザ適応部に送られる。ユーザ適応部は送られてきた検索結果の飲食店情報の特徴ベクトルを飲食店コメントコーパスから参照する。そして、その特徴ベクトルをユーザカテゴリごとに作成した特徴ベクトルを比較する。比較した結果類似度が高かったものから順に結果として表示される。ユーザは検索結果を参照するとともに、興味のある飲食店を新たな個人嗜好情報と追加できる。興味ある飲食店情報を個人嗜好情報に追加する場合は、飲食店 URI を個人嗜好情報編集部に送信し、個人嗜好情報編集部から追加登録をおこなう。

4.4. 個人嗜好情報編集部

個人嗜好情報登録部は、URI ブックマークを登録・削除・ユーザカテゴリ作成など個人嗜好情報を編集する機能を持つ。ユーザは領域オントロジーの飲食店情報の中から興味のあるもの、自分の気に入ったものを登録しておく。登録方法はあらかじめ携帯型情報端末や PC などから直接 URI を登録しておく方法と、ユーザ適応部によって出力された検索結果を用いて登録する方法がある。登録された飲食店情報はユーザの好みによって自由に分類することができ、分類するためのユーザカテゴリも自由にラベル名を付与することができる。ユーザカテゴリと登録された飲食店情報は各個人のプロフィールとして保存される。

5. おわりに

領域オントロジーと個人嗜好情報を利用することにより飲食店を対象とした携帯型情報端末用レコメン

レーションシステムを提案した。領域オントロジーはディレクトリサービスを対象とし、個人嗜好情報は携帯端末からの位置情報やカテゴリ化されたブックマーク等を用いて行った。エンドユーザが屋外で負担なく周辺情報を得られるような検索システムを目指し、ユーザの個人嗜好情報を領域オントロジーによって拡張することにより、新しい地域においてユーザの嗜好に基づいたレコメンデーションを行う試作システムを構築した。

今後は、ユーザの概念体系に基づいて分類された検索結果が領域オントロジーの概念体系においてどのように対応付けられるかを調べていくとともに、被験者を用いた実験を行い、本システムの有効性の検証を行いたいと考えている。

参 考 文 献

- [1] Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web, Scientific American, vol. 284, pp.34-43, 2001.
- [2] World Wide Web consortium: Resource Description Framework (RDF), <http://www.w3.org/RDF>
- [3] World Wide Web consortium: Web Ontology Language (OWL), <http://www.w3.org/2001/sw/WebOnt>
- [4] Gruber, T.: A translation approach to portable ontology specifications, Knowledge Acquisition, Vol. 5, No. 2, pp.199-220, 1993.
- [5] DARPA Agent Markup Language Program: DAML Ontology Library, <http://www.daml.org/ontologies>.
- [6] Swoogle, <http://pear.cs.umbc.edu/swoogle>
- [7] 伊藤 英毅, "オントロジーを利用した知識の共有/再利用", UNISYS TECHNOLOGY REVIEW 第 64 号, February . 2000 .
- [8] 佐保田 圭介, 波多野 賢治, 宮崎 純, 吉川 正俊, 植村 俊亮, "ブックマークの階層構造情報を組み込んだ協調フィルタリングによる Web ページの推薦手法", 電子情報通信学会第 15 回データ工学ワークショップ (DEWS2004), Mar .2004 .
- [9] 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継, "Bookmark からの共通話題ネットワークの発見手法の提案とその評価", 人工知能学会論文誌 17 巻 3 号 SP-D, pp . 276-284, 2002 .
- [10] 官上大輔, 河合由紀子, 田中克己, "A3: オントロジーの共有によるユーザ適応のためのフレームワークの提案", 電子情報通信学会第 15 回データ工学ワークショップ (DEWS2004), 2004 .
- [11] 大島裕明, 田中克己, "個人オントロジーを基にした Web 情報検索に関する研究", 電子情報通信学会第 15 回データ工学ワークショップ (DEWS2004), 2004 .
- [12] 仲川こころ, 高田喜朗, 関浩之, "検索目的を反映したカテゴリ構造に基づく WWW 検索支援", 情報処理学会研究報告, 2000-DBS-120, pp . 1-8, Jan . 2000 .
- [13] Yahoo! グルメ, <http://gourmet.yahoo.co.jp> .
- [14] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸: 日本語形態素解析システム『茶筌』 version 2.0 仕様説明書 第二版, NAIST Technical Report, NAIST-IS=TR99012, 1999 .
- [15] 中川裕志, 森辰則, 前田朗, 島浩之, "専門用語自動抽出システム", <http://gensen.dl.itc.u-tokyo.ac.jp> .
- [16] Salton, g., singhal, A., Buckley, C. and Mitra, M., "automatic text Decomposition Using Text segments and Text Themes", Hypertext '96 Proc., pp.53-65, April. 1996.
- [17] Gerard Salton, Michael J. McGill: Introduction to Modern Information Retrieval, McGraw-Hill, 1983.
- [18] 高野明彦, 丹羽芳樹, 西岡真吾, 今一修, 久光徹, 汎用連想計算エンジン "GETA", <http://geta.ex.nii.ac.jp>, 2002 .