

ユーザプロフィールとしてのオントロジーの構築とその評価

官上 大輔[†] 河合由起子[†] 田中 克己^{†,††}

[†] 情報通信研究機構けいはんな情報通信融合研究センター

〒619-0289 京都府相楽郡精華町光台 3-5

^{††} 京都大学大学院情報学研究科 社会情報学専攻

〒606-8501 京都市左京区吉田本町京都大学情報学研究科社会情報学専攻

E-mail: †{kanjo,yukiko}@nict.go.jp, ††tanaka@dl.kuis.kyoto-u.ac.jp

あらまし 情報をユーザの知識や嗜好に合わせて選択し提示することが可能なシステム，すなわち適応システムへの要求が強くなっている．筆者らは適応システムの構築を容易にするために，適応に必要な機能を提供するフレームワークを提案してきた．このフレームワークでは，ユーザのプロファイルはユーザオントロジーと呼ばれるオントロジーによって表現されており，ユーザとフレームワーク上の適応システムとのインタラクションに基づいて自動的に構築される．本研究では，ウェブから得られる情報を用いて，実際に多数のユーザオントロジーを構築し，提案しているフレームワークにおけるユーザオントロジーの構築手法などについて評価を行う．対象は書籍とし，ウェブから得られる個人の所有している書籍データ，および書籍の属性情報などを用いてユーザオントロジーを構築した．その結果から，書籍の各属性に関する制約を持ったカテゴリが，ユーザオントロジーの階層構造の中のどのような位置に構築されるか，といった点について知見を得ることができた．

キーワード ユーザ適応，個人適応，プロフィール，オントロジー

Studies on the Use of an Ontology as a user profile

Daisuke KANJO[†], Yukiko KAWAI[†], and Katsumi TANAKA^{†,††}

[†] Keihanna Human Info-Communication Research Center, Communications Research Laboratory,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan

^{††} Dept. of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan

E-mail: †{kanjo,yukiko}@nict.go.jp, ††tanaka@dl.kuis.kyoto-u.ac.jp

Abstract We previously proposed a framework for a user adaptation, called “Adaptation Anywhere & Anytime (A3)” [17]. In this framework, an ontology called “user ontology” is applied as a user’s profile. The user ontology is defined to be able to represent the user’s knowledge and interests, and is automatically acquired through the interaction between a user and information systems on the A3. In this paper, we discuss the user ontology and the algorithm for constructing it. To evaluate a user ontology as a user profile, we actually construct the user ontology using book data according to the proposed algorithm and then discuss it. We could have acquired some findings on user ontologies.

Key words User Adaptation, Personalization, Profile, Ontology

1. はじめに

ウェブやウェブにアクセス可能な情報機器の発達と普及とともに，ユーザが手に入れることのできる情報は著しく増加した．しかし，情報の増加は入手を容易にする一方で，適切な情報の発見を難しいものになっている．その結果，適応技術への要求がさらに強くなっている．すなわち，ウェブや個々

の情報システムに存在する大量の情報の中から，ユーザの知識や関心に応じて適切なものを選択し，提示することのできる適応性を持ったシステム（適応システム）が必要となってきた．実際，google (<http://www.google.com>) や yahoo (<http://www.yahoo.co.jp/>) などのメジャーなサイトでも，個人化可能なページを用意し，その内容をユーザが決められるようにするといった機能を備えるようになってきている．

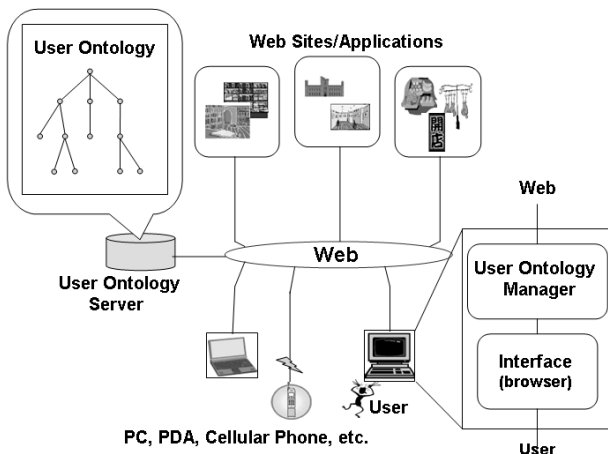


図 1 A3 の構成

適応システムの構築には様々な機能が必要となる。例えば、ユーザの知識や関心を記述したプロフィールを獲得する機能や、ユーザに提示すべき情報を当該ユーザのプロフィールに基づいて適切に選択する機能などが必要となる。しかし、プロフィールの獲得や情報選択のためのアルゴリズムを考案し、それを実装するのはシステム開発者にとって時間と手間を要する作業である。そこで我々は、適応システム構築のために必要な上記のような機能を提供するフレームワーク (Adaptation Anywhere & Anytime: A3 フレームワーク) を提案してきた[17]。A3 が提供する適応のための機能は XSLT や JSP などから利用できる。システム開発者は XSLT などの一般的なウェブサイト構築の手段を用いて、容易に適応性を持ったシステムを構築できる。

我々の提案する A3 フレームワークでは、オントロジーが重要な役割を果たす。適応を実現するためにはユーザの知識や関心を記述したユーザプロフィールが必要となる。A3 フレームワークではユーザプロフィールはオントロジーで表現されており、ユーザオントロジーと呼ばれている。ユーザオントロジーは、ユーザとシステム間のインタラクションに基づいて動的に更新される。A3 フレームワークは、ユーザオントロジーの構築機能や、ユーザオントロジーに基づき当該ユーザにとって適切な情報を選択する機能、あるいはユーザにとって既知の情報を省く機能などをシステムの開発者に提供している。

A3 フレームワークにおいてユーザオントロジーは、ユーザがシステムの利用によって獲得したと考えられる知識 (リソース) を分類した分類木として定義され、ユーザプロフィールとして利用される。この分類木であるユーザオントロジーをプロフィールとして利用する妥当性について検討する必要がある。そこで本研究では、実際にウェブから得られるデータを使ってユーザオントロジーの構築実験を行う。得られたユーザオントロジーについて、その階層構造やカテゴリの制約として用いられる属性などについて考察し、知識表現としてのユーザオントロジーとその構築手法に関する妥当性について検討する。

本稿の構成は以下の通りである。2 章で提案している A3 フレームワークについて概略を述べる。3 章ではユーザオントロジーの構築手法や、リソース選択の際のユーザオントロジー

の利用について述べる。4 章では我々が行ったユーザオントロジーの構築実験とその結果について述べる。5 章で関連研究について概観した後、6 章で本稿のまとめを記す。

2. A3 フレームワーク

本章では、我々の提案している A3 フレームワーク (Adaptation Anywhere & Anytime フレームワーク、以下、A3) について述べる。A3 は、適応システム構築のためのフレームワークであり、システムに適応性を与えるために必要な機能を提供している。図 1 に A3 の構成を示す。A3 上には、各種のシステム (Web Sites/Applications) が、A3 の提供する機能を利用して構築される。ユーザは自らの周囲にある様々な情報機器を用いて、任意の場所 (Anywhere) と時刻 (Anytime) に、A3 上の各システムを利用し、そこから自らの知識や関心の応じた適応的な出力を得ることができる。このユーザの知識や関心を表すプロフィールはユーザオントロジーによって表現され、ユーザとシステム間のインタラクションに基づいて動的に構築される。

A3 の主要な構成要素として、User Ontology Manager と User Ontology Server の 2 つが存在し、これらはユーザオントロジーの構築や更新などの操作に関する役割と、ユーザが利用する情報端末へのユーザオントロジーの配信と管理に関する役割を担う。以下、A3 の 2 つの主要な要素について述べる。

2.1 A3 の構成要素

2.1.1 User Ontology Manager

現在、A3 が提供する適応機能として、ユーザオントロジーの動的構築、ユーザオントロジーに基づく情報リソースの適応的選択の 2 つの機能がある。選択の際に既知の情報リソースを特定することも可能である。本研究では、リソースとは、実世界における事物を表す情報であり、その情報によって対象となる事物を特定できるものと定義する。具体的には、ニュースサイトのニュースやショッピングサイトの商品などである。

User Ontology Manager (UOM) は、ユーザが利用する各情報端末に存在し、上記の 2 つの機能を実際に行う主体の要素である。UOM はユーザオントロジーの構築やリソースの選択など、ユーザオントロジーに関連する全ての操作を行う。ユーザオントロジーに対する操作を全て UOM で行うことで、同じ方針に基づいてユーザオントロジーが構築されるため、一貫性が保証される。また、ユーザオントロジーは UOM と、次に述べる UOS の間を移動するのみで、A3 上に構築されたシステムがユーザオントロジーについて情報を得ることはなく、ユーザのプライバシーが保たれる。

2.1.2 User Ontology Server

User Ontology Server (UOS) はネットワークを介して UOM に接続しており、ユーザオントロジーの UOM への配信と管理の役割を担っている。ユーザが A3 上の適応システムの利用を開始すると、当該ユーザを特定し、そのユーザオントロジーをユーザが利用している端末の UOM へと送る。ユーザがシステムの利用を終えると、ユーザオントロジーを UOM から回収し、次に利用されるまでユーザオントロジーを管理する。

ユーザの利用に際して、UOS がユーザオントロジーの配信

や管理を行うことで、ユーザが利用する端末に関わらず、同じユーザに対して同じユーザオントロジーが利用されることが保証される。それに共ない、システム毎に異なるプロファイルが構築され、結果としてシステムによって適応精度が異なる、といった従来の適応システムに存在していた問題が解決される。

2.2 A3における適応過程

システムがユーザに提供できるリソースを複数持つ場合、A3の提供する適応機能を用いることで、より適切なリソースの選択ができる。また、任意のリソースをユーザへと提示し、ユーザが当該リソースに関する知識を得たと判断できる時には、ユーザオントロジーを更新できる。以下、A3におけるリソースの選択とユーザオントロジー構築の過程について示す。

A3 フレームワークでリソースの選択を行うためには、適応システムはユーザの提示可能なリソースを、一旦、リストによって表現する必要がある。リストはシステムから UOM へと送られる。UOM ではユーザオントロジーを用いて、リストの各リソースに対して重み付けを行う。重みはユーザの当該リソースに対する関心の強さを示すものであり、大きな値を持つものはよりユーザにとって関心の強いリソースとなる。より重みの大きなリソースを選ぶことで、ユーザにとってより関心の強いリソースを選択できる(3.2 節参照)。このリソースの選択機能は、ユーザの関心や興味に基づくものであり明示的な要求について考慮するものではない。ユーザの明示的な要求に反するリソースでも、システムの作成するリスト中に存在し、大きな重みが与えられた場合、ユーザへと提示される可能性がある。したがって、リストを作る時点で、ユーザからの要求を考慮する必要がある。選択されたリソースはシステムへと戻され、そこで他に必要な処理が行われる。あるいは UOM からブラウザ等のインタフェースを介して直接ユーザへと提示することできる。提示の際の形式はリストに限るものではない。

ユーザオントロジー構築の過程は以下ようになる。ユーザにリソースが提示され、ユーザがそのリソースについての知識を得たと考えられる時にユーザオントロジーは構築される。ユーザによる知識の獲得はユーザからの入力によって判断する。特定の入力ユーザからあった場合、対象となるリソースを当該ユーザのユーザオントロジーに追加することでユーザオントロジーを構築・更新する(3.1 節参照)。A3 の方では、あるシステムが利用される際にどのような入力が存在し、それがどのような意味を持つものであるかを予め知ることができない。よって、システムの開発者はどの入力ユーザによる知識の獲得を意味するものかを特定し、その入力起きた際にユーザオントロジーへと追加されるリソースについて定義する必要がある。

上記の機能を利用する手段として XSLT の拡張要素などを提供している。例えばリソースの選択のためには、リソースのリストを記した XML[23] で書かれたリストを用意し、XSLT スタイルシートで <a3:sort>要素を用いて対象リソースを特定すれば重み付けが行われ、ユーザにとってより関心のあるリソースが選択できる。このように A3 を用いることで、容易に適応システムが構築できる。

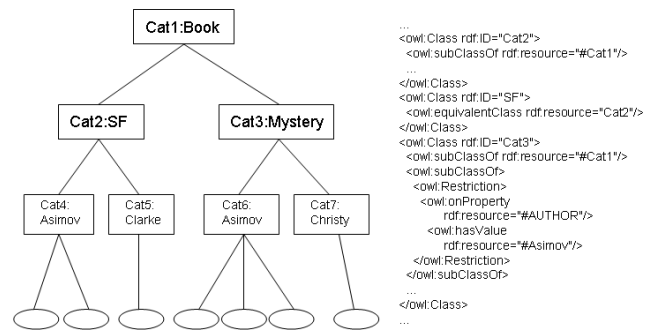


図 2 オントロジーの例

3. ユーザオントロジー

A3 では、ユーザの知識や関心を示すユーザプロファイルとして利用され、適応の過程において動的に構築される。ユーザオントロジーはリソース(図 2 左、楕円)とカテゴリ(図 2 左、四角形)よりなる分類木として定義される。リソースとは既に述べたように、ショッピングサイトにおける商品のような実際の事物を表す情報であり、全てのリソースは特定のカテゴリにカテゴリ化される。一方、カテゴリはそれぞれ固有の制約を一つ持つとともに、上位カテゴリの制約を継承している。制約はリソースに関する属性-属性値の対で、当該カテゴリにカテゴリ化されるリソースはその制約を満たす必要がある。本研究では、各リソースは属性-属性値に関するメタデータを持ち、ユーザオントロジー構築の際には、その情報を利用できると仮定している。

ユーザオントロジーとリソースは RDF/S [1],[3] と OWL [9] を用いて記述される(図 2 右側)。これらの記述法はセマンティック Web [2] で提唱されているリソースやメタデータの記述法である。これらを用いることでリソースやユーザオントロジーの再利用・共有が可能となる。また、OWL などの記述力は高く、各種の制約などを記述でき、容易に拡張をできるといった利点もある。

3.1 ユーザオントロジーの構築

2.2 節で述べたように、ユーザオントロジーはユーザが特定のリソースに関する知識を得たと判断される時に、そのリソースをユーザのユーザオントロジーへと追加することで行われる。本研究では、ユーザによる知識の獲得はユーザからの入力に基づいて判断できると仮定している。例えば、ショッピングサイトにおける商品購入を認証する入力は、ユーザがその商品を入力することを意味し、結果、その商品の知識を得ていると考えられる。このように特定の入力はユーザの知識獲得を示すものと考え、その入力があった際にユーザオントロジーを構築する。

一般に人は本を読む時や食事を食べる時、物を買う時には自らの好みや関心の強いものを選ぶ。その結果、関心の強いリソース(本、食事、物など)に関する知識を多く持つようになると考えられる。この考えに基づいてユーザオントロジーの構築は以下を仮定している。すなわち、ユーザはより関心の強いリソースに関する情報を多く獲得し、また、関心のあるリソースに共通の属性はユーザにとって重要な属性であると仮定し、リ

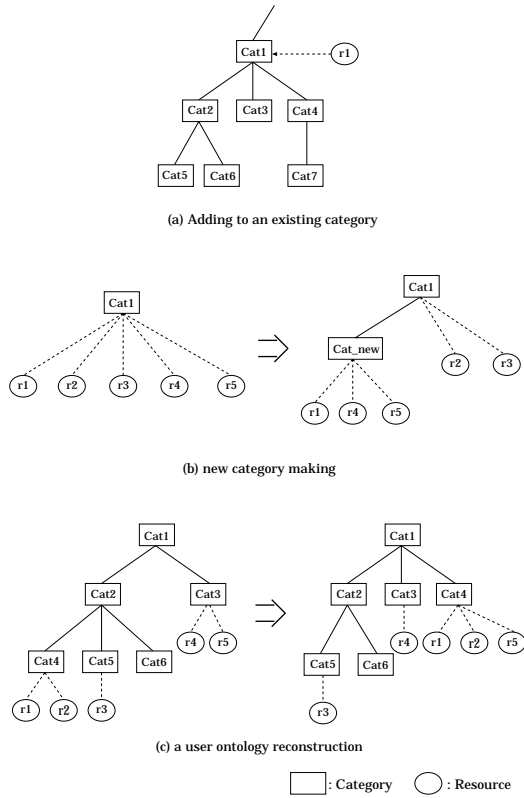


図 3 オントロジーの構築

ソースの追加やカテゴリの作成,あるいはカテゴリの移動など,ユーザオントロジーの構築が行われる.ユーザオントロジーの構築手続きを以下に示す.

a. 既存カテゴリへのカテゴリライズ

ユーザオントロジーへのリソースの追加が起きると,UOMはまず,追加されたリソースを既存カテゴリへカテゴリライズする.カテゴリライズされるカテゴリは当該リソースが制約を満たすカテゴリの中で,ルートカテゴリからの距離が最も大きなものとする.該当するカテゴリが複数ある場合,より多くのリソースを持つカテゴリにカテゴリライズされる.これは上で述べたユーザはより関心の強いリソースに関する情報を多く持つ,という仮定を反映した操作である.図3(a)は,リソース r が既存カテゴリ Cat_1 にカテゴリライズされる様子である.この例では, r は $Cat_2 \sim 7$ の制約を満たさず,もっともルートカテゴリからの距離の大きい Cat_1 へカテゴリライズされる様子を示している.

b. 新規カテゴリの作成

手続き a の結果,同じカテゴリにカテゴリライズされ,かつ,同じ属性-属性値の組を持つリソースの数が任意の閾値を越える場合,その属性-属性値の組を制約とする新しいカテゴリを作成する.また,共通の属性-属性値の組を持つリソースを新しいカテゴリにカテゴリライズする.本研究では閾値として 3 を用いている.図3(b)は,この操作によって新しいカテゴリ Cat_{new} が作られる様子を示している.リソース r_5 がカテゴリ Cat_1 にカテゴリライズされ, r_1, r_4, r_5 が共通の属性-属性値

の組を持つ時に,新たなカテゴリ Cat_{new} が作成されている様子を示している.この時, Cat_{new} の制約は r_1, r_4, r_5 に共通の属性-属性値の組となる.

c. カテゴリの移動

手続き b によって新カテゴリ Cat_{new} が作成された時,より多くのリソースをカテゴリライズできるより上位の位置へと移動する可能性がある.これはユーザはより関心の強いリソースに関する情報を多く持つという仮定と,より多くのリソースに共通する属性-属性値の組は重要である,という仮定に基づく操作である.本研究ではユーザオントロジーの上位のカテゴリを,下位のカテゴリよりも重要なものと定義している.

Cat_n の親カテゴリを Cat_{n-1} ,さらにその親カテゴリを Cat_{n-2} とする.この時, $W_{Cat_{re}}$ が $W_{Cat_{orig}}$ よりも大きくなった場合, Cat_n を Cat_{n-2} の下位カテゴリになるように移動する. $W_{Cat_{re}}$ と $W_{Cat_{orig}}$ はそれぞれ,カテゴリの移動を行った場合と行わなかった場合のユーザオントロジーに与えられる重みであり,次式で求められる.

$$\sum_{k=1}^m \left(\frac{1}{2}\right)^{x_k-1} * n_k * W_{Cat_k} * d_{tmp_k} \quad (1)$$

m : ユーザオントロジーに存在するカテゴリの数

x_k : カテゴリ Cat_k とルートカテゴリの距離

n_k : Cat_k にカテゴリライズされているリソースの数

W_{Cat_k} : n_p/n_q で与えられるカテゴリ Cat_k の重み. n_q はユーザオントロジーにおける全てのリソースの数, n_p は Cat_k の固有の制約となっている属性-属性値の組を持つリソースの数である.

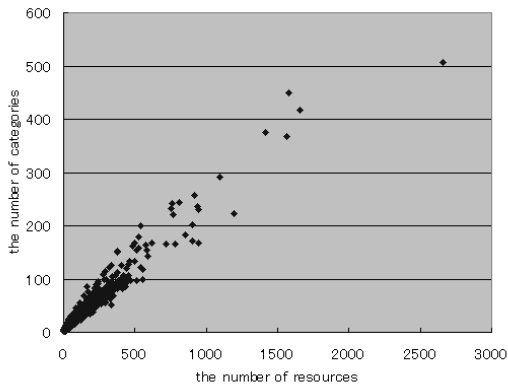
d_{tmp_k} : 最後に Cat_k にリソースがカテゴリライズされてからの期間によって決まる 0 から 1 の数値.ユーザの関心の推移を示す.この値が存在するため,ユーザオントロジーへリソースが追加される順序によって,構築されるユーザオントロジーは異なったものとなる可能性が存在する.

図3(c)にカテゴリの移動が生じる際のユーザオントロジーの様子を示す.新たに構築されたカテゴリ Cat_4 が Cat_1 の下へと移動することで,リソース r_5 などの再配置が行われ, $W_{Cat_{re}}$ が $W_{Cat_{orig}}$ よりも大きくなるため, Cat_4 の移動が成立する.

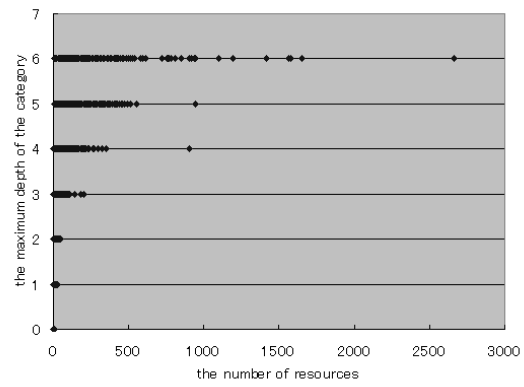
3.2 ユーザオントロジーとリソースの選択

3.1節で述べたようにユーザオントロジーは,ユーザが任意のリソースについて知識を得たと判断される時に,そのリソースをユーザオントロジーへ追加することで構築される.したがって,ユーザオントロジー中の任意のカテゴリにリソースがカテゴリライズされている場合,ユーザはそのリソースに関する知識を持っていると判断できる.既知リソースの特定はリソースの有無によって判断できる.

また,3.1節で述べたように,ユーザのリソースへの関心の強さを持っているリソースに関する知識の量が比例すると仮定



(a) Relation of the number of resources and the number of created categories



(b) Relation of the number of resources and the maximum depth of the category

図 4 リソースとクラスの関係

すると、ユーザオントロジー中のリソースの数によってユーザの関心の強さを計ることができる。すなわち、複数のリソースを考えた時、より多くのリソースがカテゴリ化されているカテゴリにカテゴリ化できるリソースの方が、そうでないリソースよりもユーザの関心は強いと考えられる。この考えに基づき、A3 ではより多くのリソースを持つカテゴリにカテゴリ化され得るリソースに大きな重みを与えることでリソース選択を行っている。対象リソースが Cat_n にカテゴリ化され、その上位カテゴリを $Cat_{n-1}, Cat_{n-2}, \dots, Cat_0$ とした場合の、重みの計算式を以下に示す。

$$\sum_{k=1}^n \left(\frac{1}{2}\right)^{k-1} * w_{r_{Cat_k}} \quad (2)$$

$w_{r_{Cat_k}}$ は各カテゴリの重みであり、 n_{k_p}/n_q で与えられる。 n_{k_p} は Cat_k 以下にカテゴリ化されているリソースの数、 n_q はユーザオントロジー中の全リソースの数である。ルートカテゴリからの距離に応じて $(\frac{1}{2})^{k-1}$ を乗じているのは、多くのリソースに共通の属性-属性値の組をより重要な属性とみなすという仮定に基づく構築操作を行っている 3.1 節 c の操作によって、より上位のカテゴリの制約となっている属性-属性値の組を重視した重みの計算をするためである

4. 実験

オントロジーの妥当性を検討するために、実際にウェブから得られるデータを用いてユーザオントロジーの構築実験を行った。本章では、その実験について記し、得られた知見について述べる。

4.1 実験手法

構築されるユーザオントロジーの一般的な傾向などを把握するため、できるだけ多くのユーザオントロジーを構築することを考えた。そのためには多くのユーザのデータが必要となる。そこで、特定の被験者に時間をかけてデータを入力してもらう手法は採らず、ウェブからデータを獲得することを考えた。

「本棚.org (<http://pitecan.com/Bookshelf/>)」は、ユーザが自分の所有する書籍を登録でき、また、他のユーザはそれを

自由に閲覧できる書籍に関するサイトである。このサイトには 2000 を越えるユーザが書籍を登録しており^(注1)、多くのユーザに関するデータを得ることができる。そこで我々はこのサイトに書籍を登録している各ユーザについて、それぞれが登録している書籍データを用いてユーザ毎に書籍を対象としたユーザオントロジーの構築を行った。構築手順を以下に示す。

1. 本棚.org に書籍を登録している各ユーザについて、登録している全ての書籍の ISBN を獲得する。ISBN を特定できない書籍については除外する。
2. 獲得した ISBN を用いて各書籍の属性-属性値の組のデータを他のウェブサイト (Yahoo!Books, <http://books.yahoo.co.jp/>) から獲得する。対象とする属性は、タイトル、作者、ジャンル、出版社、版型、価格、出版年月日の 7 つである。ウェブから適当なデータが得られなかった場合、当該の属性の属性値は空のまま扱った。
3. 各ユーザごとに、そのユーザの書籍リストにある書籍をリストの先頭から順に追加して、ユーザオントロジーを構築する。リソースの追加をリストの順序に従って行ったため、この実験では 3.1 節の (1) 式の時間に関する変数 d_{tmp_k} については値を 1 として考慮していない。

4.2 実験結果

4.1 節で述べたようにして「本棚.org」の各ユーザに対して、ユーザオントロジーの構築を行った結果、2322 人のユーザについてデータを獲得することができた。一人のユーザが登録している書籍数の最高は 2660 であった。ユーザオントロジーの構築過程におけるカテゴリ作成の閾値として 3 を用いているが、最低でも一つのカテゴリを作るために必要な 3 冊より少ない数の書籍しか登録していないユーザは 975 人 (全登録者数の 42%) であった。

図 4 は 1 人のユーザが登録している書籍の数と (a) 構築されたユーザオントロジーにおけるカテゴリの数の関係 (b) カ

(注1): 2005 年 6 月の時点。2006 年 1 月現在では 4000 を越えている。

テゴリの階層数（ルートカテゴリからの距離）の最大値との関係について示したものである。作成されるカテゴリ数が登録している書籍数におよそ比例するのに対して、カテゴリの階層数の最大値はそうでないことがグラフより分かる。

表 1 に書籍の各属性とカテゴリ数の関係について記す。ここではタイトルは一意性が高くカテゴリを構成しにくい属性として、タイトルに関する制約を持つカテゴリは省略している。表 1 の A は、各属性についてその属性-属性値の組を制約として持つカテゴリの数を示している。また、図 5 のそれぞれのグラフは、各属性に関する制約を持つカテゴリがユーザオントロジー中のどの階層に作成されているかを示したものである。属性によってカテゴリの作成される階層に傾向があることがわかる。表 1 の B は、制約となっている属性-属性値の組の異なる属性値の数を示している。

4.3 考察

4.3.1 属性に関する考察

3.1 節で述べたようにユーザオントロジーの構築は、リソースの属性-属性値の組に関する制約を持ったカテゴリを形成、あるいは移動することで行われる。本節では対象リソースとして書籍の属性と構築されたユーザオントロジーの関連性について考察する。

表 1 が示すように、出版社のようなカテゴリ数が多くとも属性値の取り得る値の種類は少ないものや、逆に出版日のようにカテゴリ数が少くとも属性値の種類は多いものが存在している。この結果からは、属性とその属性に関する制約を持ったカテゴリの数の間には関連性がないと考えられる。

一般的に書籍を分類する際に用いられる属性としてはジャンル、作者、出版社、版型などが考えられる。一般的な家庭の本棚や書店の棚、あるいは図書館などでも、書籍の分類にはこれらの属性が用いられている例が多い。一方、出版日や価格で分類している例は、古書店など一部を除き限られている。しかし、表 1 が示すように、構築されたユーザオントロジーには価格や出版日に関する制約を持つカテゴリも多い。

しかし、図 5 を見てみるとジャンル、出版社、版型といった書籍の分類に用いられることの多い属性に関する制約を持つカテゴリは、より上位のカテゴリとなっている割合が高く、一方、出版日や価格に関する制約を持つカテゴリは下位のカテゴリとなっていることが示されている。さらに版型やジャンルといった種類の少ない属性については、より上位のカテゴリとなっている様子が顕著である。これらは実際に行われる分類に用いられる属性を反映した結果であり、直観に反することのない結果を示している。

このように構築されたユーザオントロジーは実際の分類を反映した構造となっており、一般的な妥当性を持っていると考えられる。また、そのユーザオントロジーを構築している構築手

法についても同様に妥当性を持つと考えられる。

4.3.2 リソースの追加順序に関する考察

提案するユーザオントロジーの構築アルゴリズムでは、リソースをユーザオントロジーに追加する順序が影響する。そこで本棚.org に所有する書籍を登録している同一のユーザに対してユーザオントロジーを構築する際、書籍をリソースとして追加する順序を変更する、という実験を行った。その結果、ユーザオントロジーの構造の変化が大きいのに対して、構築されるカテゴリ数などに大きな変化のないことが分かった。これはユーザが持っている知識の量自体は変わらないことを示唆していると考えられる。

4.3.3 対象リソースに関する考察

本研究における対象リソースは書籍である。また、ユーザオントロジーの構築を行う際、データの収集を「本棚.org」という書籍の登録を行うウェブサイトから行っている。本節では、これらの対象リソースに関する事項について考察する。

まず、対象リソースとしての書籍について考える。これまで出版された書籍の数は膨大であり、また、今後も多くの書籍が出版されていくと思われる。ユーザの振舞いを考えると、購入や所有が容易であり、ウェブ上での検索の機会なども多い。同様の性質を持つものとしては音楽 CD や映画などがある。このようなリソースを対象とした場合、ユーザの振舞いが継続的に行われるとユーザオントロジーに加えられるリソースの数も大きなものになる。その結果、3.1 節の(1)式、 w_{cat_k} の n_p の値が大きくなる。式の有効性が失われる可能性がある一方で、一度安定した知識体系は変更されにくくなる。大量のリソースの中の 1 つのリソースの影響の大きさについて考えても、その影響が大きいと考えるに難しく、直観に反することのない結果であると考えられる。

異なったドメインの対象、例えば住居や自家用車などをリソースとした場合、また違った結果になると考えられる。これらは書籍や CD などとは異なり、絶対に満足しないといけな条件が多い、所有数や継続的な検索の可能性が少ない、といった点に相異がある。従って上記の例とは逆に w_{cat_k} の n_p が小さい状態のまま継続し、ユーザオントロジーとしての有効性が発揮できない可能性がある。したがって、有効なユーザオントロジーの構築には一定のリソース数が必要であると考えられる。

次に情報獲得源としての「本棚.org」の妥当性について考える。構築した 2322 人のユーザのユーザオントロジーの中に、属性「ジャンル」に関する制約を持つカテゴリが 5159 存在した。その中で属性値が「コミック、アニメ」であったものが 684（全カテゴリの約 13%）、「コンピュータ」であったものが 669（同 13%）存在する。例えば Yahoo!Books に登録されている全書籍は約 166 万冊であり、その中でコミックに分類されるものは約 8%（約 13 万冊）、コンピュータに分類されるものが約 3%（約 47000 冊）であることを考えると、上記の 2 つの属性を持つリソースの割合は大きいものであると考えられ、これは「本棚.org」という特定のウェブサイトの利用者の傾向を示すものと考えられる。

また、本棚.org には、所有している全ての書籍を登録しよう

表 1 各属性とカテゴリ数の関係

	作者	出版社	版型	ジャンル	出版日	価格
A	5240	8346	6081	5159	1136	7039
B	1943	437	27	19	437	543

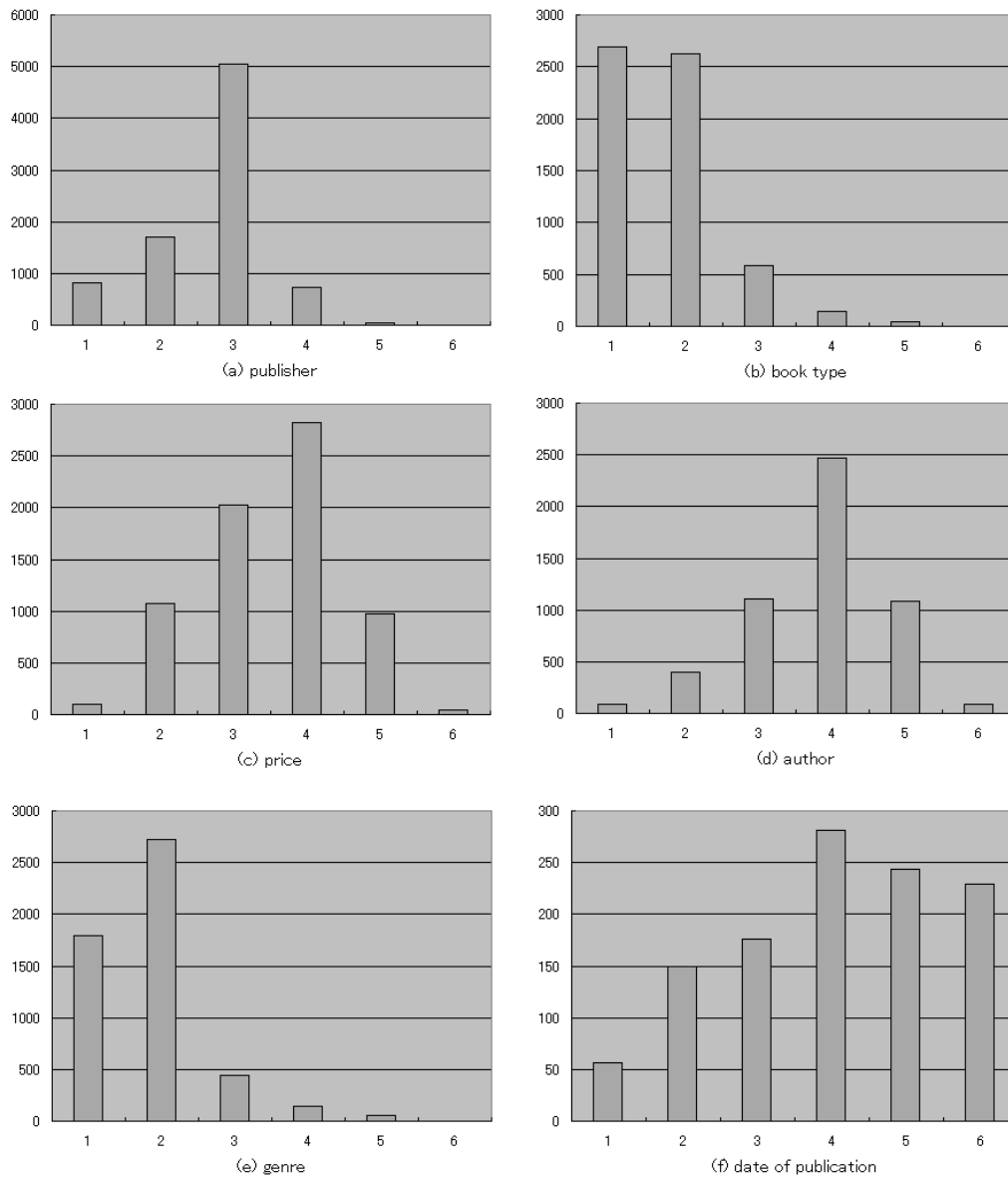


図 5 クラスの階層数と属性-属性値間の関係

とするユーザがいる一方で、「ひとまず学業に関係あるものだけ」「家をたてたい、と思った人にオススメの本をあつめてみました」というようなユーザのコメントにあるように特定のテーマを持って登録するユーザもいる。このようなデータを用いて構築されるユーザオントロジーは、「本棚.org を利用する」というコンテキストにおけるユーザオントロジーであり、当該ユーザの一般的・普遍的なものとは異なるものと考えられる。ただし、A3 フレームワークを用いた場合、他のサイトの利用の際にも同一のユーザオントロジーが用いられるため、このような特定サイトにおける偏向を解消できる可能性を持っている。

5. 関連研究

一部のウェブサイトでは、個人化の機能が提供されている。例えば Yahoo! (<http://www.yahoo.co.jp/>) や google (<http://www.google.co.jp/>) では、提示されるウェブページのレイアウトやそのコンテンツをユーザ毎に決定できる機能が提

供されている。しかし、これらのページではユーザが自らコンテンツ等を明示的に定義する必要があり、ユーザのプロファイルといった観点からの知見を得られるものではない。

セマンティックウェブ [2] は、計算機に意味的な処理を行わせることを目的として、ウェブ上の様々なコンテンツにメタデータを付与する試みである。本研究でユーザオントロジーの記述に用いている RDF [1], [3] や OWL [9] は、セマンティックウェブにおけるコンテンツやリソース、オントロジーの記述法の一つとして提案されている XML [23] ベースの記述法である。

オントロジーは多様な目的のために利用されており、様々な観点から研究されている。まず、オントロジーそのものの作成や構築を目的として、複数のオントロジーの間でマッピングやマージを行うための手法に関する研究がある [13], [14], [21]。しかし、これらの研究は手法について述べたものであり、実際に構築したオントロジーについて論じたものではない。一方、我々は実際に提案する構築アルゴリズムでオントロジーを作成した。

その結果に基づいて考察を行い、構築されたオントロジーと構築手法について妥当性を検証している。

また、ユーザモデリングに関する研究としては [4], [12], [16], [19], [20], [22] などがある。これらの中にはセマンティック Web の技術を用いたものも存在する。これらはそれぞれに興味深い結果を出している。しかし、その多くが特定のアプリケーションなどに強く関連付けられており、実際にシステムの構築を考えるものがこれらのユーザモデル、あるいはモデルの構築手法を使う際に、どのように利用すれば良いかが分からないものが多い。他にもデータクラスタリングや分類木の学習など、本研究と関連する研究が数多く存在する。しかし、それらのユーザモデリングや、あるいはデータクラスタリングの手法を使ってユーザモデルを構築し、さらに利用するためのアルゴリズムや機構を考案・開発する手間の多くはシステム開発者に残されている。一方、A3 は XSLT などを用いた構築手法を提供しており、適応システムとユーザモデルの切り分けを明確にしている。ただし、今回はユーザオントロジーや構築手法についての妥当性に主眼を置いたため、個人のプロファイルとしての検証は不十分であり、今後、この観点からの検討は必要である。

6. おわりに

我々は適応システムの構築を支援するため、ユーザ適応を実現するのに必要な機能を提供するためのフレームワークである A3 (Adaptation Anywhere & Anytime) を提案している。このフレームワークでは、ユーザプロファイルとしてオントロジー (ユーザオントロジー) が用いられる。ユーザオントロジーはユーザが知識を得たと判断される際に動的に構築される。ユーザは関心を持つものについてより多くの知識を持つとすると、ユーザオントロジーからは、ユーザが持っている知識と関心についての情報を得ることができる。

本研究では、ウェブから得られる情報を用いて、A3 の中で実際に利用されているユーザオントロジーの構築アルゴリズムに従って実際にユーザオントロジーを構築した。その結果、ユーザオントロジーの階層構造とリソースの属性の関係性などについての知見を得ることができた。その結果、ユーザオントロジーの構造は、一般的な分類を反映したものであり、人の直観に反するものではないことが確認できた。

今後は他のドメインのリソースを用いた構築実験を行い、今回、書籍を対象にして得られた結果と比較などを行い、オントロジーの妥当性について、さらに検討する必要があると考えられる。また、ユーザオントロジー構築のためのデータの収集方法などについても考える必要がある。また、今回の結果からはユーザオントロジーの一般的な妥当性を検証することができたが、個人プロファイルに関する検討は不十分である。今後は実際にユーザオントロジーを用いた適応性についての検証を行う必要がある。

文 献

- [1] Beckett,D.,RDF/SML Syntax Specification, <http://www.w3c.org/TR/rdf-syntax-grammar/>
- [2] Berners-Lee,T.,Hendler,J.,Lassile,O., The Semantic Web, Scientific American(2001)
- [3] Brickley,D.,Guha,R.V., RDF Vocabulary Description Language 1.0:RDF Schema, <http://www.w3.org/TR/rdf-schema/>
- [4] Chafee,J.,Gauch,S., Personal Ontologies for Web Navigation, In Proc. of CIKM'00(2000)
- [5] Chan,H.,Finin,T., An ontology for context-aware pervasive computing environments, Special Issue on Ontologies for Distributed Systems, Knowledge Engineering Review (2003)
- [6] Chen,C.,Chen,M.,Sun,Y.,PVA:A Self-Adaptive Personal View Agent, Journal of Intelligent Information System(2002)
- [7] Chen,L.,Sycara,K., A personal agent for browsing and searching, In Proc. 2nd Intl. Conf. Autonomous Agents(1998)
- [8] Clark,J., XSL Transformations(XSLT) Version1.0, <http://www.w3.org/TR/xslt>, 1999
- [9] Dean,M., Connolly,D., Harmelen,F., Hendler,J., Horrocks,I., McGuinness,D., Schneider,P., Stein,L., OWL Web Ontology Language 1.0 Reference, <http://w3.org/TR/owl-ref/>
- [10] De Bra, P., Stash, N., De Lange, B., AHA! Adding Adaptive Behavior to Websites. Proceedings of the NLUUG Conference(2003)
- [11] De Bra, P., Aerts, A., Berden, B., De Lange, B., Rousseau, B., Santic, T., Smits, D., Stash, N., AHA! The Adaptive Hypermedia Architecture. Proceedings of the ACM Hypertext Conference(2003)
- [12] Dolog,P.,Nejdl,W., Challenges and Benefits of the Semantic Web for User Modeling, Adaptive Hypermedia and Adaptive Web-Based System AH2003(2003)
- [13] Fridman,N.,Muse,M.A., An algorithm for merging and aligning ontologies:automation and tool support, In Proc. of the Workshop on Ontology Management at the Sixteenth National Conference on Artificial Intelligence(AAAI-99), AAAI Press(1999)
- [14] Fridman,N.,Muse,M.A., SMART:Automated Support for Ontology Merging and Alignment, In 12th Workshop on Knowledge Acquisition, Modeling and Management(1999)
- [15] Grimnes,G.A.A.,Chalmers,S.,Edwards,P.,Preece,A., 'Granitenights - a multi-agent visit scheduler utilizing semantic web technology ' Seventh International Workshop on Cooperative Information Agents(2003)
- [16] Henze,N.,Nejdl,W., Knowledge Modeling for Open Adaptive Hypermedia, In Proc. of Adaptive Hypermedia and Adaptive Web-Based System AH2002(2002)
- [17] Kanjo,D.,Kawai,Y.,Tanaka,K., Ontology based Framework for Adaptive Web System, Workshop I:Enterprise Modeling and Ontology: Ingredients for Interoperability, in Intl. Conf. of Practical Aspect of Knowledge Management(2004)
- [18] Kay,J.,Kummerfeld,B.,Lauder,P., Personis: A Server for User Models, In Proc. of Adaptive Hypermedia and Adaptive Web-Based System AH2002,pp203-212(2002)
- [19] Kobsa,A., Generic User Modeling Systems, User Modeling and User-Adapted Interaction 11,pp49-63(2001)
- [20] Kurki,T.,Jokela,S.,Sulonen,R.,Turpeinen,M., Agents in delivering personalized content based on semantic metadata, In Proc. 1999 AAAI Spring Symposium Workshop on Intelligent Agents in Cyberspace,pp84-93(1999)
- [21] Noy,N.F.,Musen,M.A., The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies, 59(6),2003
- [22] Pretschner,A., Gauch,S., Ontology Based Personalized Search, In Proc. 11th Intl. Conf. on Tools with Artificial Intelligence (1999)
- [23] Extensible Markup Language (XML), <http://www.w3c.org/XML/>
- [24] Tanudjaja F., Mui L., Persona: A contextualized and personalized Web search, In Proceedings of the 35 Annual Hawaii International Conference on System Sciences(2002)