

Peer-to-Peer 環境におけるユーザ間のインタラクションに基づく階層的知識の拡張手法

佐藤 宏之^{†,††} 杉本 重雄^{††}

情報リソースのメタデータの記述に利用する共通のボキャブラリの定義やオントロジを Web 上に置いて、複数のユーザや組織の間で参照するための研究が進められている。本論文は、個別のユーザや組織によって生成され分散環境上に公開された階層的知識をユーザ間で参照して活用する方法について述べる。ここではオブジェクト指向の考えを利用して、情報リソースをインスタンスとしたときのクラスに相当する概念を定義し、概念を階層化して表現した階層的知識を扱っている。セマンティック Web のフレームワークを利用して、概念と結びついた情報リソースに関するメタデータを記述し、それを Peer-to-Peer 環境で流通させることによって、他者が保持する階層的知識の発見を支援する方式を提案する。さらに、ユーザの「関わった情報を整理して保持したい」というモチベーションを利用して、自らのために行った活動から他者とのインタラクションを発生させ階層的知識を拡張するプロセスモデルを提案する。このプロセスを支援するシステムを用いた実験では、ユーザは自らが生成した階層的知識に他者の概念を引用したり、他者の概念との間の関係付けを行ったりして、階層的知識を拡張することが可能であることを確認した。また、拡張の際に強く関わりあう 3~6 人のユーザの階層的知識を結合することで、ユーザ間で平均約 71% の概念の引用や関係付けが類似するなどの特徴を持った新たな階層的知識の生成が確認できた。

Evolution of Concept Hierarchies through Interaction among Users in a Peer-to-Peer Environment

HIROYUKI SATO^{†,††} and SHIGEO SUGIMOTO^{††}

Many studies that have the goal of providing multiple users and organizations with Web-based common vocabularies or ontologies for use in describing metadata on informational resources are now under way. This paper describes a method that facilitates reference to and utilization of concept hierarchies evolved by individual users or organizations and made available in distributed environments. In the method we use an object-oriented paradigm to define resource-related concepts as classes, the instances of which represent resources. We set up these classes in a hierarchy that represents the hierarchy of concepts. We use the Semantic Web as a framework for a way to define metadata that reflects linkages between concepts and resources in a way that makes it easy for users to find associated concept hierarchies held by other users in a peer-to-peer environment. Moreover, we propose a model for the evolution of concept hierarchies through interaction among users. This process takes advantage of the motivation of users to organize information which they find relevant. We implemented the system and ran an experiment which demonstrating that hierarchies created by using the system evolve as the users refer to or make connections with concept hierarchies held by other users. We also identify an interesting feature of new concept hierarchies produced by applying the evolutionary process to merge concept hierarchies initially produced by sets of three to six "strongly related" users: actions common to the users make up an average of 71% of all actions (connection and reference) that lead to the evolution of new hierarchies.

1. はじめに

World Wide Web (Web) 環境の特徴は「オープン」かつ「分散」である。この特徴を活かして情報リソースを共有するだけでなく、情報リソースのメタデータや、メタデータの記述に利用する共通のボキャブラリの定義やオントロジを複数のユーザや組織の間で共有して活用するための研究が進められている。

† 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
NTT Information Sharing Platform Laboratories, NTT Corporation

†† 筑波大学大学院図書館情報メディア研究科
Graduate School of Library, Information and Media Studies, University of Tsukuba

情報リソースの意味を記述するためのデータモデルとして Resource Description Framework (RDF)¹⁾がある。これは特定のアプリケーションを前提としない知識表現方法を規定している。RDFの最小のデータモデルはリソース、プロパティ、値の三つ組によって構成される。たとえば、情報リソースとしてあるドキュメントがあったとする。そのプロパティとして「著者」、値として「山田太郎」を指定した場合、ドキュメントの著者は山田太郎であるということを曖昧性なく表現できる。このため、RDFは機械による処理に適した汎用のメタデータ記述言語として利用が可能である。

セマンティック Web²⁾では、さらにオブジェクト指向の考えを利用し、情報リソースをインスタンスとしたときのクラスに相当する概念を定義し、概念の階層構造や概念間の制約などの関係を記述する仕組みを提供している。これらを Web 上でオープンに参照できるようにすることで、機械が情報の意味を自動処理できるようにすることを目的としている。

本論文では、個別のユーザや組織によって生成された概念の階層構造を階層的知識にとらえ、分散環境上に公開された階層的知識をユーザ間で参照して活用する方法について述べる。セマンティック Web のフレームワークを利用して、概念と結びついた情報リソースに関するメタデータを記述し、それを Peer-to-Peer 環境で流通させることによって、他者が保持する階層的知識の発見を支援する方式を提案する。さらに、ユーザが自身のために行った活動から他者とのインタラクションを発生させ階層的知識を拡張するプロセスモデルを提案する。またユーザがそのプロセスに従った活動を行うことを支援するシステムの説明とそれを用いた実験結果を報告する。

2. セマンティック Web における階層的知識

Web 上でオントロジを扱う試みは、すでに 1995 年から SHOE (Simple HTML Ontology Extensions) プロジェクト³⁾において実践されている。現在では、World Wide Web Consortium (W3C)⁴⁾によって、RDF を基盤に RDF Vocabulary Description Language (RDF Schema)⁵⁾や、米国の DAML プロジェクトで開発されたオントロジ記述言語 DARPA Agent Markup Language Ontology notations (DAML-ONT)⁶⁾とヨーロッパの On-to-Knowledge プロジェクトで開発された Ontology Inference Layer (OIL)⁷⁾を統合した言語である DAML+OIL をベースとした Web Ontology Language (OWL)⁸⁾が W3C 勧告と

なり、標準化が完了している。W3C はこれらを、セマンティック Web の構想の実現に必要な段階の 1 つと位置付けている⁹⁾。

セマンティック Web におけるオントロジ記述言語は概念間の同等性や非重複性などの関係や概念に対する制約を記述できるなどの高度な表現力を持つが、重要な表現力の 1 つに概念間の階層関係表現がある。RDF Schema は情報リソースをインスタンスとしたときのクラスに相当する概念を定義し、概念間の階層関係を定義することができる。また、OWL も概念間の階層関係の定義には RDF Schema を利用する。

RDF Schema や OWL は、ユーザや組織によって記述された概念定義を Web 上に公開して活用することを意図して開発されている。概念に Uniform Resource Identifier (URI) を与え、Extensible Markup Language (XML) の形式で概念定義を記述すれば Web 上に公開できる。それを第三者が参照し、別の場所で定義されている概念との間のマッピングを新たに記述することもできる。

セマンティック Web では、Upper オントロジ^{10),11)}のように最初から統一された唯一の概念体系を定義するというよりは、ドメインやコミュニティに依存したオントロジが個々に生成され、定義の一部などがユーザや組織間で互いに引用されて拡張されることを指向しているといえる。引用にともない概念の階層関係も拡張されると考えられる。

3. 分散環境における階層的知識拡張の問題

オントロジはあるユーザや組織の知識が特定の知識表現を用いて明示化されたものといえる。オントロジに記述された概念階層をユーザや組織の階層的知識と位置付けた場合、これは以下のように活用できる。

たとえば、ある組織では学会の研究会発表で利用した資料を、「ドキュメント」という概念の下位概念に「プレゼン資料」という概念を定義して、これと結びつけて整理していたとする。この階層的知識は、まずは組織内で情報を分類したり検索したりする際に活用できる。さらにこの「ドキュメント」と他のユーザが定義した「文書」という概念が同等であるということを知ると「文書」という概念のみを知っている第三者に対しても下位概念として「プレゼン資料」を提示することができるようになる。また、その概念に結び付けられた資料を検索対象とすることができると考えられる。

「ドキュメント」と「プレゼン資料」の関係のように、ある組織などにおいて活動するユーザ自らが定義

した概念階層は、そのドメインやコミュニティにある知識を最も反映することができると考えられる。しかし、個々に記述された階層的知識は、そのままでは、そのユーザや組織に閉じた範囲でのみ通用するものになりがちである。これらを他者が参照したときに「他人が記述した概念階層は使えない」という問題に直面する。文献 12) では、カールスルーエ大学の実験において被験者は他者によって用意された概念階層を用いて作業することは非常に難しいという結果を示している。

この問題を解決する方法の 1 つに、上記の「ドキュメント」と「文書」の関係のように、他のユーザや組織との間で概念間の関係を構成することが考えられる。各ユーザや組織が定義した概念と、他者の概念との相対的な関係を明らかにすることで階層的知識の利用を促進するのである。しかし、階層的知識の構成においては、概念化の対象とした情報リソースの種類や概念階層の構築者によって異なった階層構造を有する¹³⁾ため、簡易に複数の階層的知識を統合して拡張することはできないという問題がある。

4. 他者の概念を取り込んだ階層的知識拡張

4.1 問題解決のためのアプローチ

本研究では、他者の階層的知識を利用するには、他者の複数の階層的知識の中から自らの階層的知識に関係のある概念を適切に探し出し、概念間の関係付けを簡易に行い、階層的知識を拡張することを支援する仕組みを提供することが重要であると考えられる。階層的知識を構成する個々の概念定義だけを参照しても他者との共通点を探すことは難しい。そこで、概念と結びついた情報リソースに関するメタデータを記述し、それを Peer-to-Peer 環境で流通することによって、分散環境下で他者が保持する階層的知識の発見を支援する方式を提案する。さらに、ユーザの「関わった情報を整理して保持したい」というモチベーションを利用して、自らのために行った他者の概念の引用や自ら定義した概念と他者の概念との関係付けなどのユーザ間のインタラクションを発生させ、階層的知識を拡張するプロセスモデルを提案する。

4.2 本研究で扱うメタデータと階層的知識

概念は RDF Schema や OWL などのオブジェクト指向を取り入れた知識表現の観点からみると、インスタンスをともなったクラスとして扱うことができる。本研究では、このクラス間の階層関係や同等関係を「階層的知識」としている。図 1 に本研究で扱うクラスとインスタンス、さらにインスタンス間の関係を例

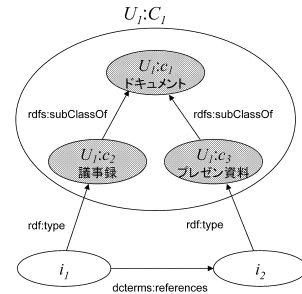


図 1 メタデータと階層的知識のグラフ表現

Fig. 1 Graph representation of metadata and concept hierarchy.

示した。 $U_1 : C_1$ という記号で示されたユーザ U_1 の階層的知識の中にそれぞれ $U_1 : c_1$, $U_1 : c_2$, $U_1 : c_3$ という記号が「ドキュメント」、「議事録」、「プレゼン資料」という 3 つの概念がクラスとして存在し、「ドキュメント」を親クラスとした階層関係を構成している。「議事録」と「プレゼン資料」を表すノードから `rdfs:subClassOf` というラベルが付けられた有向のアーキが「ドキュメント」とつながり、グラフ構造を構成している。

クラスのインスタンスとなるものとしては、Web ページや Web 上に置かれた研究会発表で利用したスライド資料などの URI を持つ電子的なコンテンツを想定している。本研究ではこのインスタンスから他のノードへつながるアーキが形成するラベル付きの有向グラフで表現されるデータを「メタデータ」としている。たとえば、図 1 において、インスタンスを表すノードである i_2 は「プレゼン資料」を表すノードに対して `rdf:type` というラベルが付けられた有向のアーキがつながり、クラスとインスタンスの関係を示している。

さらに、 i_1 と i_2 が `dcterms:References` というラベルを付けたアーキで結ばれているように、インスタンスどうしでもグラフ構造を構成することがある。このインスタンス間の関係を示す RDF のプロパティとしてはさまざまなボキャブラリの利用が可能である。たとえば情報リソース全般に関する基本的な共通語彙は Dublin Core Metadata Initiative (DCMI) によって、Dublin Core Metadata Element Set (DCMES)^{14),15)}として公開されている。ここでは“Title”や“Creator”などの 15 の基本となる要素が標準化されている。さらにより厳密なメタデータ記述のために、DCMI は限定子 (DC Qualifier) を定義している。たとえば情報リソースの関係を示す基本要素である“Relation”は“references”, “isReferencedBy” (参照・被参照),

“hasVersion”, “isVersionOf” (異なるバージョン), “requires”, “isRequiredBy” (要件) などのボキャブ ラリを用いて詳細記述することが可能である .

4.3 プロセスモデル

次に本研究で提案するプロセスモデルの詳細を 3 つのフェーズに分けて説明する . 本プロセスでは , まず各ユーザに「自身のために関わった情報を整理して保持したい」というモチベーションがあることを前提に考 えている . Web ブラウザのブックマークやコンピュ タのファイルシステムのディレクトリなどは , ユーザ によって階層的にコンテンツが整理されている場合が 多い . このように自身が関わった情報を階層構造を利 用して整理することはユーザにとって自然なことであ ると考えられる .

4.3.1 フェーズ 1 : 各ユーザごとのメタデータと 階層的知識生成

階層的知識を構成するにあたって , 何も無いところ からユーザに構成させることは難しいと考えられる . そこで , 最初のフェーズではユーザが扱うコンテンツ 間を関係付ける作業を支援したり , コンテンツの集合 に対してその共通概念をクラスとして付与したりする ことを支援する . 生成された複数のクラスを整理する 段階で必要に応じて階層構造を生成して整理すること になる .

このフェーズ 1 において , すでに図 1 に例示したよ うな , ユーザ U_1 によるメタデータと階層的知識が生成される .

4.3.2 フェーズ 2 : 他のユーザ知識の参照

フェーズ 2 ではユーザは他のユーザの保持するコン テンツとそのメタデータを参照する . ユーザは自身が フェーズ 1 で利用したコンテンツに関して , 他のユーザは「どのコンテンツと関係付けを行っているのか」, また「どのクラスのインスタンスとしているか」を , フェーズ 1 で他のユーザが生成したグラフをたどることによって参照する .

これはコンテンツに他のユーザからも一意に参照で きる URI が与えられていることにより可能となる . 同 じ URI を持つコンテンツを利用したユーザ間の階層 的知識は図 2 に示すように共通の URI を示すノード によって結合できる . ここでは他のユーザ U_2 が持つ クラスやインスタンスを表すノード間のアークは破線 で示している .

共通の URI に関する複数のユーザのメタデータを Peer-to-Peer プロトコルを用いて収集することにより , 複数箇所に分散したユーザの階層的知識を動的にたど ることが可能になる .

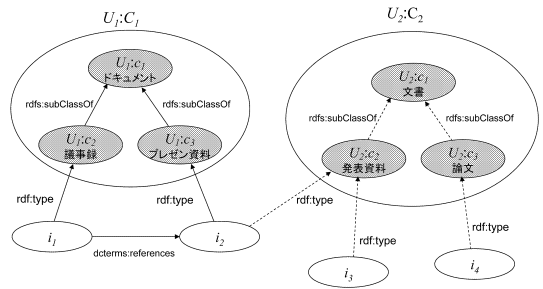


図 2 フェーズ 1 で生成されたデータ (左) とフェーズ 2 で参照可 能な他ユーザのメタデータと階層的知識 (右)

Fig. 2 Metadata and concept hierarchies produced by users U_1 and U_2 (right) in phase 1. User U_1 is able to refer to the latter in phase 2.

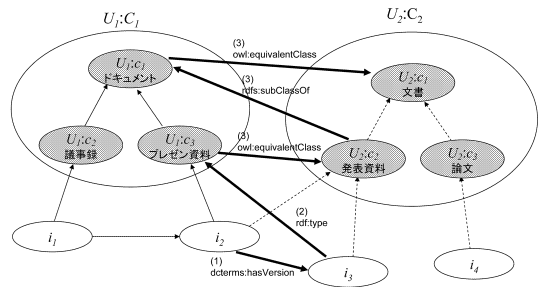


図 3 フェーズ 3 における他ユーザデータの引用と関係付け Fig. 3 Quotation from and linkage with data held by another user in phase 3.

4.3.3 フェーズ 3 : 階層的知識の拡張

フェーズ 3 では , フェーズ 2 での参照をふまえて , 他ユーザのコンテンツや階層的知識と自身のクラス などを関係付ける . ここでは以下の種類の関係付けが ある .

- (1) 他ユーザが保持するコンテンツと自身の保持するコンテンツを結ぶ関係付け
- (2) 他ユーザが保持するコンテンツを自身のクラスのインスタンスとする関係付け
- (3) 他ユーザのクラスと自身のクラスを結ぶ関係付け (他ユーザのクラスを自身のクラスのサブクラスとして引用したり同等クラスとしたりするなど)

図 3 にこれらの関係付けの様子をグラフ構造で示す . ユーザ U_1 が U_2 のメタデータや階層的知識を参照して自身の階層的知識 C_1 や自身が保持するコンテンツとの間に新たに生成可能なアークの例が太線で示されている . 初期状態からフェーズ 1 に移行した後は , 各ユーザは非同期に参照や階層的知識の拡張を繰り返 し , フェーズ 1 からフェーズ 3 のいずれかの状態にあることになる .

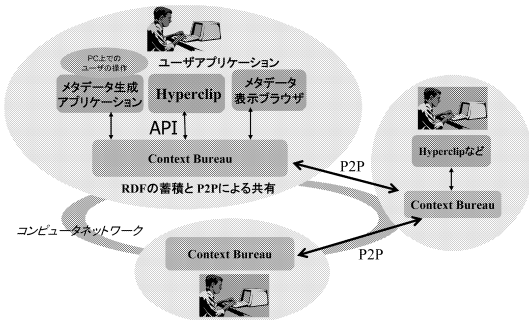


図 4 コンテキストビューロを利用したシステム

Fig. 4 Context Bureau as part of a wider system.

5. 階層的知識拡張を支援するシステム

上記のプロセスを支援するためのシステムとして、コンテキストビューロ (Context Bureau)^{16),17)} とそれに接続して動作するアプリケーションソフトウェアの 1 つである Hyperclip を利用した。

5.1 支援システム

5.1.1 コンテキストビューロ

コンテキストビューロはユーザが日常的に利用するコンピュータ上に常駐して、ユーザが扱うコンテンツに関するメタデータなど RDF で表現可能なデータを管理するアプリケーションソフトウェアである。図 4 にコンテキストビューロを中心としたシステムアーキテクチャを示した。同じコンピュータ上の他のアプリケーションソフトウェアと通信するための API を持ち、メタデータを受け取って蓄積したり、検索要求に対して指定された条件にあったメタデータを RDF の形式で返したりする機能を持つ。メタデータに相当するデータと RDF Schema や OWL で記述されたクラス定義に関するデータの両方を扱うことができる。さらにアプリケーションソフトウェアからの要求に応じて、ネットワーク上の複数のリモートコンピュータ上で動作するコンテキストビューロに蓄積された RDF データを Peer-to-Peer (P2P) のプロトコルを用いて検索できる。

1 つのコンテキストビューロに対して複数のアプリケーションソフトウェアが接続可能である。そのため、ユーザのアプリケーションソフトウェア上での操作などから、コンテンツ間の関係や属性情報を取得してメタデータを生成し、コンテキストビューロに送るアプリケーションソフトウェアや、コンテキストビューロにコンテンツに関連するメタデータを検索させ、その結果を視覚的に表示するアプリケーションソフトウェア¹⁷⁾ を独立して動作させることも可能である。

5.1.2 Hyperclip

Hyperclip はコンテキストビューロに接続して動作するアプリケーションソフトウェアの 1 つである。これはユーザのコンピュータのデスクトップ上に常駐し、ユーザが扱うさまざまなタイプのローカルファイルやメール、ブラウジング中の Web ページなどのコンテンツを、ブックマークのようにドラッグ&ドロップで登録できるツールである。本研究ではコンピュータのデスクトップ上でアイコンとして表現されたコンテンツを、ドラッグ&ドロップで Hyperclip 上で管理されているコンテンツやクラスと関係付けることを「クリップする」と呼んでいる。図 5 に Hyperclip のユーザインタフェースを示す。Hyperclip はメタデータエディタとしての機能も有する。

5.2 システムによるユーザのプロセス遂行の支援

コンテキストビューロと Hyperclip を利用したユーザは提案したプロセスを以下のように遂行することができる。

5.2.1 フェーズ 1 の支援：自ら発見したコンテンツのクリップ

図 5 の「ユーザ登録情報表示エリア」は左から順番に、1 人のユーザが自身で利用するために登録したクラスとそのインスタンスであるコンテンツ、さらにコンテンツのプロパティが表示されている「議事録」クラスのインスタンスとして「検討会議事メモ」というコンテンツがクリップされており、dcterms:References という関係付けで「プレゼンテーション 1」というコンテンツがクリップされている。

ユーザはプロパティのダブルクリックによってコンテンツ間の関係の再編集ができる。また、コンテンツタイトルのダブルクリックによって別のアプリケーションソフトウェアを起動して、コンテンツの実体を参照することもできる。新たなコンテンツのアイコンをドラッグしてすでにクリップされている別のコンテンツに重ねると、コンテンツ間の関係の入力を促すプルダウンメニューが表示される。ここで、Dublin Core において定義されている RDF のプロパティなどを選択できる。ここでは 4.3.1 項のフェーズ 1 のユーザの活動を支援し、ユーザに自由にグラフ構造を持つデータを生成することを可能にしている。生成されたデータはユーザのコンピュータ上のコンテキストビューロに蓄積される。

5.2.2 フェーズ 2 の支援：他ユーザのデータ参照

図 5 の「他ユーザ登録情報参照エリア」では、ネットワーク上のリモートコンピュータ上で他のユーザが登録したクラスやコンテンツを参照できる。自身が保



図 5 Hyperclip のユーザインタフェース

Fig. 5 User interface of Hyperclip.

持するクラスやコンテンツにマウスポインタを合わせて、マウスの右ボタンをクリックしてメニューから検索を選択するだけで、その対象の URI を含んだリモートコンピュータ上に蓄積されている他者のデータを、コンテキストビューアの機能を利用して Peer-to-Peer プロトコルにより検索して表示することができる。関連するデータが多い場合、プロパティやプロパティ値に関する条件を「検索条件設定エリア」に指定することによって、結果の絞り込みができる。これらの機能は 4.3.2 項のフェーズ 2 の他のユーザの知識を参照するための活動を支援する。

5.2.3 フェーズ 3 の支援：データ参照を基にした階層的知識拡張

検索結果として表示されている他者のクラスやコンテンツはドラッグ&ドロップで自身の「ユーザ登録情報表示エリア」にクリップすることが可能である。この場合、ドラッグするものとクリップ先の対象に応じて、コンテンツを自身が持つクラスのインスタンスとしたり、他者のクラスを自身が持つクラスと同等クラスやサブクラスとしたりして関係付けることができる。これらの機能は自身の階層的知識を拡張する 4.3.3 項のフェーズ 3 のユーザ活動を支援する。

また、リモートコンピュータ上に登録されているコンテンツを自身のクラスのインスタンスとして登録すると、利用したコンテンツを保持するユーザに対してそれを通知する機能がある。これを参照したユーザが、自身のコンテンツがどこに登録されたか再度検索を行ったりすることで、フェーズ 2 とフェーズ 3 を繰り返す

返すユーザ活動が活性化すると考えられる。フェーズ 3 の結果生成されたデータも、生成を行ったユーザのコンピュータ上のコンテキストビューアに蓄積される。

6. 実験

コンテキストビューアと Hyperclip を利用して、本研究で提案する他者の知識を利用した階層的知識拡張のプロセスが有効に動作するか、またどのように階層的知識が拡張するかを観察するために実験を行った。

6.1 実験内容

12 人の被験者に対して、Web 上の複数のコンテンツを提示し、これらの間の関係付けやコンテンツをインスタンスとしたときのクラスの生成、さらにクラス間の階層関係などを生成してもらった。生成したものは今後自らが利用し続けることを想定して作業してもらい、作業終了後にアンケートにも回答してもらった。コンテンツは、検索エンジンを利用して得られた海外のある島の観光情報に関する検索結果から 50 個収集したものであるが、ポータルサイトのトップページなどのように複数の情報が 1 つのページに集約されているものは除いている。

6.1.1 課題 1

各被験者は 50 個の Web コンテンツの中からランダムに選ばれた 30 個のコンテンツを参照し、Hyperclip のフェーズ 1 を支援する機能を利用して、コンテンツ間の関係付けや、ユーザ自らがクラス階層を作成しながら、そのコンテンツをインスタンスとして登録してもらった。

6.1.2 課題 2

次に、Hyperclip のフェーズ 2 を支援する機能を用いて、自らが登録したコンテンツをキーに他のユーザが関係付けを行っているコンテンツやそのコンテンツをインスタンスとしている他のユーザのクラスを参照してもらった。さらにフェーズ 3 の支援機能を用いて、発見した他のコンテンツやクラスを引用したり、自らが保持していたコンテンツやクラスとの関係付けを行ったりして、コンテンツ間の関係構造やクラス階層を変化させてもらった。

なお課題 2 は本来、同期もしくはランダムな順番で非同期的に行われるべきである。ユーザによっては、課題 2 によって他ユーザに拡張された概念階層をさらに参照して課題を遂行することになる。しかし、本実験ではこの影響を取り除くため、各ユーザの課題 1 を終了した状態のメタデータや階層的知識を保持した。12 人の被験者が、それぞれ課題 2 の拡張が行われていない状態の課題 1 の結果のみを参照した。

6.2 実験結果

6.2.1 クラスに関する操作回数

表 1 は、課題 1 と 2 において各ユーザがクラスに関してどれだけ操作を行ったかを示す。他ユーザクラス引用数は、ユーザが他ユーザのクラスを自ら保持するクラスのサブクラスや親クラスとして引用した数である。他ユーザクラスとの関係登録数については、自分のクラスと同等であるとユーザが考えた他ユーザクラスとの間を owl:equivalentClass プロパティで結んだ数である。

6.2.2 アンケート結果

課題 2 終了後のアンケートでは、「後からコンテンツを参照しやすくするために引用したり関連付けたりしたい他者のクラスはあったか?」という問いに対しては全ユーザ (100%) が「はい」を選択した。また、「他者のクラスを参照することによってクラス階層は良くなったか?」という問いに対しては、実際に課題 2 で階層を変化させた 70% のユーザが「はい」を選択した。その理由として、「いろいろなまとめ方を簡単に引用可能」、「より細かい分類ができた」、「自分が登録していない階層があったため」といった回答があった。最後に、他者の概念階層をただ参照するのと比べて、今回の実験のように「コンテンツをキーにコンテンツと結びついた他者のクラスを参照することで他者クラスを引用したり関連付けたりしやすくなったか?」という設問に対しては約 92% のユーザが「はい」を選択した。

表 1 各課題におけるユーザごとのクラスに関する操作回数
Table 1 Number and type (recording in phase I, quotation or linkage in phase II) of user actions related to class.

ユーザ	課題 1 のクラス登録数	課題 2 の他ユーザクラス引用数	課題 2 の他ユーザクラスとの関係登録数
uA	6	9	1
uB	17	7	16
uC	7	15	31
uD	14	2	51
uE	4	0	9
uF	23	1	42
uG	12	8	0
uH	21	0	33
uI	30	14	46
uJ	8	5	1
uK	11	5	25
uL	14	5	25
全体	167	71	287

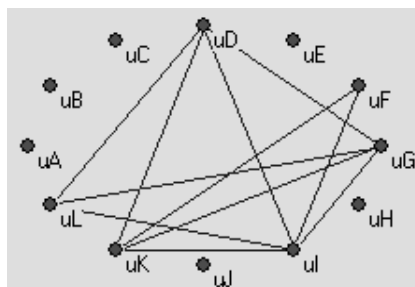


図 6 ユーザ I (uI) と強い関連性があるユーザによって構成されたユーザ間のネットワーク

Fig. 6 Network between user I (uI) and users having strong relationships with uI.

6.2.3 ネットワーク分析の結果

各ユーザのクラス階層のうち、クラスを頻繁に引用しあっているなど他ユーザのクラスとの結びつきが強いユーザ間の階層的知識をマージして分析した。ユーザ間の結び付きの強さは、2 者間で少なくともどちらか一方のユーザが、もう一方のユーザの 2 つ以上のクラスを引用しているか、5 つ以上のクラスとの関係付けを行っている場合に強い関連性があると判断した。なお、複数のユーザ間に形成されたコンテンツ間の関係付けについては、ユーザごとに初期コンテンツはランダムに与えられておりコンテンツの初期の重なりにはばらつきがあることから、関連性の強さを判断する材料としていない。

各ユーザごとに、保持するクラスの被利用状況から強い関連性を持ったユーザを抽出し、さらに抽出されたユーザ間で関連性が強い関係を抽出した。図 6 は各ユーザのノードを表示し、ユーザ I (uI) のクラスを利用したユーザのうち強い関連性を持ったユーザ間

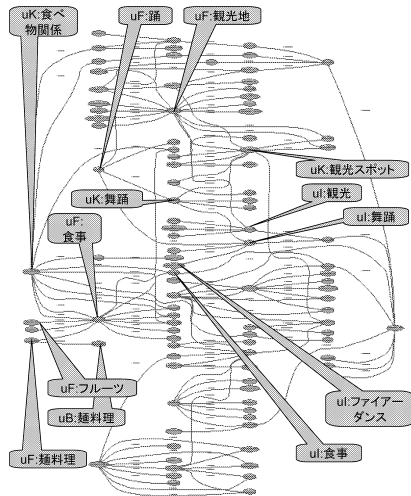


図 7 階層的知識集合 {C_{I1}, C_{F2}, C_{K2}} 内のクラスをノードとしたグラフ表示

Fig.7 Graph representation of concept hierarchy set {C_{I1}, C_{F2}, C_{K2}}; each node represents a class defined by one user.

のアークを表示したものである。

次にこれらのネットワークの中から各ノードを i , それにつながるアーク数を n_i , ノード数を n としたときに, 次の式でネットワーク密度 (*density*) が 0.8 以上という条件を満たすサブグラフを抽出した。

$$density = \frac{\sum_{i=1}^n n_i}{n(n-1)}$$

ネットワーク密度が 1 の場合は, そのサブグラフは完全グラフであることを意味する。なお, 複数のサブグラフが抽出された場合, 他のサブグラフの一部を構成するサブグラフを独立したものとせず, そのサブグラフを含んだノード数 n がより大きいサブグラフのみを抽出した。この結果, 12 個のサブグラフが抽出された。これらのサブグラフを構成するユーザが保持する階層的知識をマージした結果を分析した。

図 7 は uI の課題 1 の結果生成された階層的知識 C_{I1} と, uI とサブグラフを構成する uF, uK の課題 2 の結果生成された階層的知識 C_{F2}, C_{K2} から構成した階層的知識集合 {C_{I1}, C_{F2}, C_{K2}} の各クラスをノードとしたグラフである。横方向の矢印の多くは左方向にサブクラスをたどっている。縦方向の多くはそれぞれの階層間を同等クラスを示すプロパティ (owl:equivalentClass) で結んだ関係を示している。

図 8 は同じ階層的知識集合に対して, 同等なクラスを示すプロパティで結ばれたクラスどうしを同じノードとして扱い, 同等の関係を表すアークを消して再度グラフを描画したものである。これにより, ラベルは

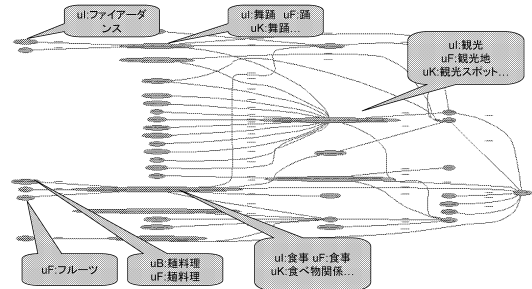


図 8 階層的知識集合 {C_{I1}, C_{F2}, C_{K2}} 内の同等の関係にあるクラスを同一のノードとしたグラフ表示

Fig.8 Graph representation of concept hierarchy set {C_{I1}, C_{F2}, C_{K2}}; nodes representing equivalent classes in the previous graph have been merged.

異なっているが同じノードとして表現されたクラス間の階層関係が形成されている。

最後にこのクラス階層と uI が課題 2 で生成した階層的知識 C_{I2} の比較を行った。uI は課題 2 で他ユーザのクラスをサブクラスとしたり, 同等としたりする関係付けを 60 回行っている。このうちの 37 回の操作で生成された関係は, uI が操作を行わなくても集合を構成する他のユーザによって uI が自身のクラスとの間に記述した関係と同じ, もしくはグラフのアークを自身のクラスノードから 2 回たどる範囲内に関係付けられていた。

表 2 では, 抽出した 12 の階層的知識の集合に対して同様の比較を行った結果を示している。ここでは, 1 人のユーザから見た場合, ユーザが自身の活動によって引用や関係付けを行ったクラス間の関係の平均約 71% は自身が活動をしなくても他のユーザによって同等もしくは類似の関係が生成されることを示している。

7. 考 察

7.1 実験結果に関する考察

実験結果に示すように, 各ユーザは分散環境において, 他のユーザの階層的知識を参照し, その中のクラスを自らのクラスに引用したり, 関係付けたりすることができている。本論文で提案したプロセスは機能していると考えられる。また, アンケート結果が示すように, 70%のユーザは他者の概念を利用することで自らのクラス階層が良くなったと回答している。また, クラスのインスタンスであるコンテンツをキーとした検索にともなうクラスの参照によって他者の概念の利用は可能という仮説に沿った行動がみられている。

一方, クラスに関する操作回数を示した表 1 のとおり, 2 人のユーザが他ユーザクラスの引用を行わず, 1 人のユーザが他ユーザクラスとの関係付けを行って

表 2 抽出された階層的知識集合と 1 人のユーザから見た階層的知識拡張の類似度
Table 2 Sets of extracted concept hierarchies and similarity of evolution from one user's point of view.

階層的知識集合	ネットワーク密度	課題 2 の操作数 (操作ユーザ)	類似の 操作数	類似度
{ $C_{B1}, C_{C2}, C_{F2}, C_{I2}$ }	0.83	23 (uB)	19	82.6%
{ $C_{B1}, C_{D2}, C_{F2}, C_{I2}$ }	0.83	23 (uB)	13	56.5%
{ $C_{C1}, C_{H2}, C_{I2}, C_{L2}$ }	0.83	46 (uC)	40	87.0%
{ $C_{D1}, C_{A2}, C_{G2}, C_{I2}, C_{K2}$ }	0.8	53 (uD)	24	45.3%
{ $C_{F1}, C_{B2}, C_{C2}, C_{I2}, C_{K2}$ }	0.8	43 (uF)	39	90.7%
{ C_{G1}, C_{A2}, C_{D2} }	1	8 (uG)	4	50.0%
{ $C_{G1}, C_{C2}, C_{D2}, C_{K2}, C_{L2}, C_{I2}$ }	0.8	8 (uG)	7	87.5%
{ C_{H1}, C_{B2}, C_{D2} }	1	33 (uH)	18	54.5%
{ $C_{H1}, C_{C2}, C_{D2}, C_{G2}, C_{K2}, C_{L2}$ }	0.8	33 (uH)	27	81.8%
{ $C_{I1}, C_{D2}, C_{G2}, C_{K2}, C_{L2}$ }	0.9	60 (uI)	41	68.3%
{ C_{I1}, C_{F2}, C_{K2} }	1	60 (uI)	37	61.7%
{ $C_{K1}, C_{D2}, C_{F2}, C_{I2}$ }	0.83	37 (uK)	34	91.9%

いない。アンケートの自由回答欄や実験後のヒアリングから、課題 2 で引用を行わず、階層構造を変化させなかったそれぞれの理由として「自分が保持する階層はシンプルなものを維持したい」、課題 1 で細かいところまで階層構造を作成したため他ユーザのクラスの引用を必要としなかった」ということが明らかになった。この 2 人のユーザは「コンテンツをキーにコンテンツと結びついた他者のクラスを検索することでクラスを引用したり関連付けたりしやすくなったか?」というアンケートには「はい」と答えており、他のクラスとの関係付けは行っている。また、他ユーザのクラスの引用のみで関係付けを行わなかった 1 人のユーザも「同じコンテンツでも人によってさまざまな整理の仕方をしていることが分かって参考になった」という理由から、同じアンケートの回答は「はい」と答えている。実験では生成したデータを実験後も自らが利用し続けることを想定してもらっており、引用や関係付けを行わない方が良いと判断したユーザには無理に操作を行ってもらっていない。引用と関係付けという手段を与えても必ずしも両方をユーザが実施するとは限らないが、実施しなかったユーザもその手段そのものには有効性を感じていると判断できる。

ネットワーク分析では、階層的知識の集合を形成した。これは自分の概念をよく参照してくれるコミュニティをユーザごとに抽出したとらえることもできる。このコミュニティとユーザの行動の類似度が大きい場合、関わったユーザ間で合意がとれれば階層的知識集合を手直して、低コストでコミュニティで共有できる共通の階層的知識を生成することができると考えられる。

しかし、同等なクラスを同じノードとして扱った階層的知識集合では、集合を構成する階層的知識が多く

なるとノードの個数が減る傾向がある。たとえば、あるユーザがクラス間の階層関係を定義していても、その両方を別のユーザが同一の関係付けをした場合、それは 1 つのノードとなっている。その結果、同じノードをサブクラスとする不適当な関係が多くみられることになる。試しにすべての階層的知識の集合からグラフを生成したところ、同じノードをサブクラスとする関係が 6 個存在した。今回は、なるべく個々のユーザが定義した階層関係を集合内で活用できるように、集合を形成するユーザ間のネットワーク密度として 0.8 以上という条件を設定している。ユーザ間で強い共通性があると思われる階層的知識集合の構成条件であったが、抽出した各集合の中に不適当な関係は平均 1.5 個存在していた。また、各ユーザ間のクラスの引用数や関係登録数を基にユーザ間の関連の強さを判断したが、結果的に関係の総登録数が多かったユーザの階層的知識が集合に多く含まれている。今回は 1 人でも任意のクラス間を同等と関係付けるユーザがいると、それが活かされたので、階層的知識の集合ではこれらのユーザの影響が大きいといえる。

今後は、不適当な関係を減らし、関係付けの個数は少なくともコミュニティにとって有益な階層的知識を提供するユーザの操作を反映できるように、ネットワーク密度のほかにも関係付けの操作数が特に多いユーザとそうでないユーザでは、ユーザ間の関連性の強弱の判断に異なる閾値を設けるなど、新たな階層的知識の抽出には、これらのパラメータの設定を検討する必要がある。

7.2 応用に関する考察

本研究で行った実験では、複数の Web コンテンツを情報リソースとした際に、リソースの内容を表す主題としてクラスが存在することが多く観察された。主

題とはたとえば情報リソースの著者などが著作中で表現したいと考えている内容のことである。主に主題は情報リソースの分類という用途に用いられるが、分類に利用する場合には、同じ意味の内容を1つの表現に統一したりするなど、統制されたボキャブラリを生成することが必要である。また、これらが階層化されているとさらに分類に利用しやすい。本研究で提案した手法は、大きな領域の体系的な知識を必要とするところではなく、特化された領域に関する主題の統制されたボキャブラリである階層的知識をコミュニティの中で作り上げるところに適用した場合に、特に効果があると考えられる。

たとえば、実験とほぼ同じ状況である「複数人である島に関する観光情報を Web 上から収集して共有する」という作業は、旅行会社の企画部門や、趣味の旅行サークルなど、10~20人程度のそれほど大きくないコミュニティ内でよく起こりうる状況である。このような状況を効率良く支援する枠組みは Web 上の情報を扱う分野で求められているものといえる。この作業において、Web コンテンツをカテゴリ分けしてリンク集に相当するものを生成したり、共有フォルダに階層化して情報を整理したりする場合、既存の方法では特定のユーザが生成した階層を中心に、階層構造を拡張するなど恣意的にせざるをえない。そのため、各個人がそれぞれ考えている階層的知識を活用することが難しいということがある。

これに対して、本論文で提案したネットワーク分析の結果発見できる、強い関連性を持った複数のユーザの階層的知識集合を利用することで、個々のユーザが自由に形成した階層的知識から類似性を持った階層が抽出できる。これを自身の階層的知識と他のユーザの階層的知識の間にある知識として共有することで、分類のための統制されたボキャブラリとして活用できると考えられる。たとえば図8に示した階層的知識集合においては、「観光」、「観光地」、「観光スポット」といった複数のユーザが抽出した主題が1つになったノードが生成されるなどの統制がなされ、階層が構成されている。これはコミュニティの階層的知識として、少し手直しするだけでコミュニティの外部に公開することなどが可能になると考えられる。RDF や OWL といった W3C で標準化されている知識表現のフレームワークを利用して記述されているため、異なるセマンティック Web のアプリケーションなどからの再活用が期待できる。

また、共有可能な階層的知識集合の各クラスは個々のユーザが生成した階層的知識のクラスへのリンクも

保持している。そのため、Hyperclip のようなユーザインタフェースを利用して個別のユーザの階層的知識をたどることができる。共有の階層的知識を通して他ユーザにおける各クラスの相対的な位置付けの違いを理解し、それらを参照することが可能になると考えられる。企業のナレッジマネジメントにおいては、他者が構築した知識をいかに発見・再活用するかが課題となっているが、この課題の解決にも寄与するものと考えられる。

8. 議 論

これまでの研究には、階層的な知識を含んだオントロジを人間が協調して編集する環境に関する研究がある¹⁸⁾。しかし、本論文では、皆が共有のオントロジを作ろうとして行動しているわけではなく、自身のために行うプロセスの中で他者の階層知識を利用したり、他者と共有できる階層知識が生成できるところに違いがあると考えている。

文献 19) の研究ではフォルダ推薦の有効性を述べているが、これはインスタンスをともなったクラスを推薦することによって、他の人に概念を提示できることを示しているのとらえることができる。その点で本研究と同じような事実を確認していると考えられる。しかし、本研究では他のユーザに概念を提示するだけでなく、概念間をユーザの知識を引き出しながら結びつけるためのフレームワークの提案に踏み込んでいる。

階層的知識の生成に関しては、人間の関わりを最小限として、概念に付けられたラベルの語彙間の近さやインスタンスの類似性に着目したオントロジの結合を行う研究がこれまでに多数ある^{13),20)}。本論文の手法ではメタデータの流通によってユーザによる概念の発見を支援するが、類似する概念の発見にはこれらの手法を組み合わせると活用できると考える。

本研究では、Peer-to-Peer 環境を利用したオープンな場所に個々の階層的知識を置きそれらを自由に検索対象とすることで、参照、引用などのユーザ活動による階層的知識の拡張を確認した。RDF Site Summary (RSS)²¹⁾ や FOAF²²⁾ などのメタデータは分散された Web サーバ上に公開されているが、これを収集するには URI を知らなければならない。しかし Peer-to-Peer 環境ではクラス定義も含めて必要な情報を動的に探索することが可能であり、アドホックな知識交換のコミュニティを形成し、その中で知識を明示化して再活用するフレームワークとしても活用できると考えられる。

また、今回はオントロジの構築手法について知識が

ないユーザもそのユーザが考える範囲で上位・下位概念を気をつけてもらうだけで、他者の階層的知識のメリットを享受できるモデルを考えたい。しかし、ここで生成された階層的知識はそのままでは RDF Schema の仕様に規定されている推移律を満たさないなど、機械における推論などの処理に耐えられるものではない。現時点では、機械が推論できるものを作るには、オントロジの作者を絞り、一貫性をもって記述することが不可欠であるという議論が多くある¹⁸⁾。人々がクラスを引用しあう際に、矛盾を考慮し制約などを継承した形で、どこまで高度なオントロジの再構築ができるかは今後の研究が必要である。

最後に、知識表現に Topic Maps²³⁾ を用いた場合にも本研究で提案する手法の適用が可能であると考えたい。今回の知識表現では、Web ページのコンテンツを概念のインスタンスとして扱っている。この場合概念となるものが、その Web ページの主題に相当するものになっていることが多いということは 7.2 節で述べたが、このようなモデル化は RDF Schema や OWL におけるクラスとインスタンスの関係だけでなく、Topic Maps におけるトピックとオカレンスの関係でとらえてみると表現がしやすい場合があると考えられる。

9. おわりに

本論文では、個別のユーザや組織によって生成された Web 上に公開された階層的な概念定義を活用するための方法として、概念のインスタンスとなる情報リソースのメタデータをセマンティック Web のフレームワークを利用して記述し、Peer-to-Peer 環境で流通することにより、関連する他者概念の発見を支援する方式を考えたい。さらに、「自身のために関わった情報を整理して保持したい」というユーザのモチベーションを利用して、自身のために行った活動から他者とのインタラクションを発生させ階層的知識を拡張するプロセスモデルを提案したい。

セマンティック Web の現状を見ると、W3C によって分散環境で利用するための知識表現記述のフレームワークは完成したが、それらを機能させる方法については研究が進められているところである。今後は、本研究で提案したプロセスを応用することで、Web 上に RDF Schema や OWL で定義された任意のドメインのポキャプラリー定義やオントロジに対して、階層的知識拡張を行うことを検討したい。

参考文献

- 1) Beckett, D.: RDF/XML Syntax Specification (Revised).
<http://www.w3.org/TR/rdf-syntax-grammar/>
- 2) Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web, *Scientific American* (2001). <http://www.sciam.com/2001/0501issue/0501berners-lee.html>
- 3) Hefin, J. and Hendler, J.: Searching the Web with SHOE, *Artificial Intelligence for Web Search, AAAI Workshop*, pp.35-40, AAAI Press (2000).
- 4) W3C: World Wide Web Consortium.
<http://www.w3.org/>
- 5) Brickley, D. and Guha, R.: RDF Vocabulary Description Language 1.0: RDF Schema.
<http://www.w3.org/TR/rdf-schema/>
- 6) Stein, L.A., Connolly, D. and McGuinness, D.: DAML-ONT Initial Release. <http://www.daml.org/2000/10/daml-ont.html>
- 7) OIL initiative: Welcome to the OIL-Page.
<http://www.ontoknowledge.org/oil/>
- 8) McGuinness, D.L. and van Harmelen, F.: OWL Web Ontology Language Overview.
<http://www.w3.org/TR/owl-features/>
- 9) W3C: W3C Semantic Web.
<http://www.w3.org/2001/sw/>
- 10) Cycorp: OpenCyc.org.
<http://www.opencyc.org/>
- 11) IEEE: Standard Upper Ontology Working Group (SUO WG) Home Page.
<http://suo.ieee.org/>
- 12) Maedche, A.: *Ontology Learning for the Semantic Web*, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers (2002).
- 13) 市瀬龍太郎, 武田英明, 本位田真一: 階層的知識間の調整規則の学習, 人工知能学会論文誌, Vol.17, No.3, pp.230-238 (2002).
- 14) 杉本重雄: Dublin Core について (第 1 回), 情報管理, Vol.45, No.4, pp.241-254 (2002).
- 15) 杉本重雄: Dublin Core について (第 2 回), 情報管理, Vol.45, No.5, pp.321-335 (2002).
- 16) 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.42, No.7, pp.742-750 (2002).
- 17) 佐藤宏之: “賢い Web” を作るための新技術「セマンティック Web 入門」第 5 回 開発者向けツールを使いこなす ネット上で配布されているメタ情報の生成/解析ツールを活用, 日経インターネットソリューション 8 月号, 日経 BP 社 (2003).
- 18) 砂川英一, 古崎晃司, 来村徳信, 溝口理一郎:

「法造」におけるオントロジー分散開発, 人工知能学会第3回セマンティックウェブとオントロジー研究会資料 (2002).

- 19) 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, Vol.17, No.3, pp.276-284 (2002).
- 20) Agrawal, R. and Srikant, R.: On Integrating Catalogs, *Proc. 10th International World Wide Web Conference*, ACM, pp.603-612 (2001).
- 21) Beged-Dov, G., Brickley, D., Dornfest, R., Davis, I., Dodds, L., Eisenzopf, J., Galbraith, D., Guha, R., MacLeod, K., Miller, E., Swartz, A. and van der Vlist, E.: RDF Site Summary (RSS) 1.0.
<http://web.resource.org/rss/1.0/spec>
- 22) FOAF project: the friend of a friend (foaf) project. <http://www.foaf-project.org/>
- 23) Garshol, L. M.: Metadata? Thesauri? Taxonomies? Topic Maps!. <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>

(平成 16 年 5 月 24 日受付)

(平成 16 年 11 月 1 日採録)



佐藤 宏之 (正会員)

平成 6 年図書館情報大学院卒業。平成 8 年慶應義塾大学大学院政策・メディア研究科修士課程修了。同年日本電信電話株式会社入社。以来、協調学習支援システム等ネットワークコミュニティの研究を行う。現在、情報流通プラットフォーム研究所にて、RDF を利用したメタデータ流通アーキテクチャおよびその応用サービスの研究開発に従事。ICCE98 Outstanding Paper Award 受賞。INTAP セマンティック Web 委員会委員。W3C RDF Data Access Working Group メンバ。電子情報通信学会会員。



杉本 重雄 (正会員)

京都大学工学部情報工学科卒業, 同大学院工学研究科情報工学専攻修了。京都大学工学博士。京都大学工学部助手を経て, 1983 年図書館情報大学院に転任。2002 年 10 月筑波大学との大学統合により, 筑波大学・図書館情報学系・教授。知的コミュニティ基盤研究センター勤務。デジタルライブラリ全般に関心を持つが, 研究の中心はメタデータに関する技術。Dublin Core Metadata Initiative では評議委員会, 諮問委員会のメンバ。