

概念共有ユビキタスドメインにおける運用センサデータを用いたオントロジ構築支援フレームワーク

三玉 政喜[†] 富井 尚志[‡]

[†]横浜国立大学 大学院環境情報学府 情報メディア環境学専攻

[‡]横浜国立大学 大学院環境情報研究院

〒240-8501 横浜市保土ヶ谷区常盤台 79-7

E-mail: [†] d07hc039@ynu.ac.jp, [‡] tommy@ynu.ac.jp

あらまし 近年、センサ類の技術的進歩やネットワークの普及などから空間の情報を取得する「ユビキタス環境」が発展していくと考えられる。このような「ユビキタス環境」では、大量のデータが取得できるが、センサから得られた生データ(raw-data)であるために、ユーザへの有意な支援やフィードバックは困難である。そこで Semantic Web で用いられるオントロジを、センサデータと組み合わせることで、raw-dataに「意味」を紐づけすることが可能となる。そこで本論文では、ユビキタスドメインによって得られる運用センサデータを用いるオントロジ構築支援フレームワークを提案する。このフレームワークを我々が提案・運用する「概念共有環境 CONSENT」に適用し、運用実験を行い、その有用性・実現性を示した。

キーワード ユビキタスコンピューティング, 時空間 DB, オントロジ

Ontology Construction Framework with Raw Sensor-data in Concept Sharing Ubiquitous Domain

Masaki MITAMA[†] and Takashi TOMII[‡]

[†] Department of Information Media and Environment Sciences, Graduate School of
Environment and Information Sciences, Yokohama National University

[‡] Graduate School of Environment and Information Sciences, Yokohama National University

79-7 Tokiwadai, Hodogaya-ku, Yokohama 240-8501 Japan

E-mail: [†] d07hc039@ynu.ac.jp, [‡] tommy@ynu.ac.jp

Abstract In recent years, ubiquitous environments in which users can share and use information that overcome from the real world became realistic. In such environment, they can get many data, but only raw sensor-data can't give them meaningful support or feedback. Then, we can give raw-data semantics by putting Ontology together with sensor-data. In this paper, we purpose Ontology construction framework that uses raw-sensor data. This framework recommends Ontology editor new Ontology that is made by sensor-data of users' action semi-automatically. We show this framework's useable and realizable by life experiment in CONSENT.

Keyword Ubiquitous Computing, Spatial-temporal DB, Ontology

1. はじめに

近年、安価で高性能な計算機やネットワークの流通や、RFIDや加速度センサのようなセンサ類の技術的進歩などから生活空間内にセンサ類を配置し、その情報を取得する「ユビキタス環境」が発展していくと考えられる。

このような「ユビキタス環境」では、大量のデータが取得できるが、生データ(raw-data)のままではユーザへの有意な支援やフィードバックは困難である。ユビキタス環境、とくにセンサネットワークでの大量のデータを削減するという手法として、Tiny DB[1]やMIST[2]などがあるが、データの意味情報がきちんと紐づけされていないとデータの扱いは困難である。さらにセンサネットワークのデータには

不完全性があるので、そのデータをDBに蓄積する際にも注意が必要で、Amol DeshpandeらはそのためにBBQという確率モデルを用いた手法[3]を提案している。このようにユビキタス環境でのデータには「大量」「不完全」「raw」といった問題が存在する。

一方で、同様の事象はすでにWebで問題となっていた。「情報爆発」という言葉で表現されるように、情報が大量に存在するWebにおいて、必要な情報の取得が困難になりつつある。そこでその解決策のひとつとしてオントロジを用いることが提案された。オントロジは人の持つ知識をマシンリーダブルな形で表現することにより、情報の詳細な論理表現を行うツールである[4]。しかしオントロジは単純

なシソーラスレベルでは普及しているものの、その構築コストやクリティカルなキラーアプリケーションといった面で問題となっている。そこで、オントロジとユビキタス環境を組み合わせることで互いの欠点を補完する研究が行われている。代表的なユビキタスとオントロジのメリット・デメリットを表1に挙げる。

これらの背景を基に我々も、実空間を RFID や加速度センサなどによってユビキタス化し、その取得データをオントロジによって意味付けすることで、高度なユーザ支援を行う「概念共有環境 CONSENT」を提案・実装し、実運用してきた。

しかし、前述したようにオントロジ構築は一般的に非常に手間がかかる。特に多くの場面で共通な常識(Upper Ontology)は共有することで構築の手間の削減ができるが、ドメインに特化した知識(Domain-Specific Ontology)はドメイン固有のものなので共有することが難しい[5]。さらにこの Domain-Specific Ontology こそが、ユーザ支援などを行う際にはもっともよく参照され、重要度も高い。よって、この Domain-Specific Ontology の構築の手間を削減する必要がある。

そこで本論文では、ユビキタス環境から得られるセンサデータを用いて、オントロジ構築の手間を削減するフレームワークを提案する。このフレームワークではユーザに特別な負担をかけずに、ユーザの行動を RFID や加速度センサから取得し、その情報を用いて、階層構造であるオントロジの類似ノードの情報などを用いてオントロジを半自動生成し、オントロジ編集者に推奨する。このフレームワークでは、オントロジ編集者が過去に記述したオントロジを用いるトップダウン的な手法と、ユーザの実際の行動のセンサデータを用いるボトムアップ的な手法の二つを提案する。以上のフレームワークを実装し、運用実験を通して取得したデータを用い、Recall/Precision による推奨精度の評価、及び作業時間などを評価しその有用性・実現性を示す。

以下、2章において概念共有環境 CONSENT を概説し、3章で本論文の提案手法について詳述する。そして5章でそれらに対する評価実験を行い、最後に6章でまとめと今後の課題について述べる。

2. 概念共有環境 CONSENT

本章では、我々がこれまでに提案してきたユビキタス環境 DB の一つである「概念共有環境 CONSENT(CONcept Sharing ENvironment)」(以下、単に CONSENT と表記する)の特徴やスキーマ・思想について概説し、このスキーマで

表1 オントロジのメリット・デメリット

	ユビキタス	オントロジ
メリット	<ul style="list-style-type: none"> ● ユーザの意識なくデータを取得 ● 行動支援などのアプリケーション 	<ul style="list-style-type: none"> ● Semantic Web による標準化 ● 意味をマシンリーダブルに記述可能
デメリット	<ul style="list-style-type: none"> ● 大量データ ● 動的データ ● 意味情報を含まない単純なセンサデータ 	<ul style="list-style-type: none"> ● 構築コスト ● キラーアプリケーション ● 深い意味の記述はまだ困難

のオントロジ構築に関するコスト(手間)の定義やその削減について考察する。

2.1. 基本モデル

ユビキタス環境と意味情報を組み合わせたソリューションである CONSENT では、実空間に存在するオブジェクトや人がオブジェクトに対して行う操作、さらに空間に存在する知識やルールなどの意味情報をオントロジとして構築している。さらに空間情報やセンサデータなどの実データであるマルチメディアデータを独立に定義し、意味情報とマルチメディアデータの間に関連付けるエンティティ(「存在エンティティ」と定義)を用いて実空間を表現し、DB に蓄積する三層構造モデル化手法[6, 7]を提案してきた。三層とは“意味情報”, “マルチメディアデータ”, “存在エンティティ”を指す。これらを独立に管理することにより、マルチメディアデータが表現する感覚的な情報のみならず意味情報をも共有することが可能となり、知識や常識などの空間の概念や実空間の現在の状態である現実状態の共有を実現した。図1に概念共有環境 CONSENT の論理スキーマを示し、以下“意味情報”“マルチメディアデータ”“存在エンティティ”を格納するテーブル群について詳述する。

2.1.1 意味情報

CONSENT の意味情報は RDF (Resource Description Framework) スキーマを基礎として構築する。RDF とは Semantic Web の基本層を形成するデータモデルであり、主語、述語、目的語の三つ組(トリプル)を基本単位として、トリプルの集合でウェブのセマンティクスを表現する[4]。一般的にトリプルの主語、目的語はクラスと呼ばれるタイプに属し、述語はプロパティと呼ばれるタイプに属する。CONSENT では図2のようにオントロジを表現する。クラスやプロパティの定義を記述する枠組みが RDF スキーマであり、CONSENT の意味情報を格納するテーブル群のスキーマに相当する。具体的には、RDF でのクラスがテ

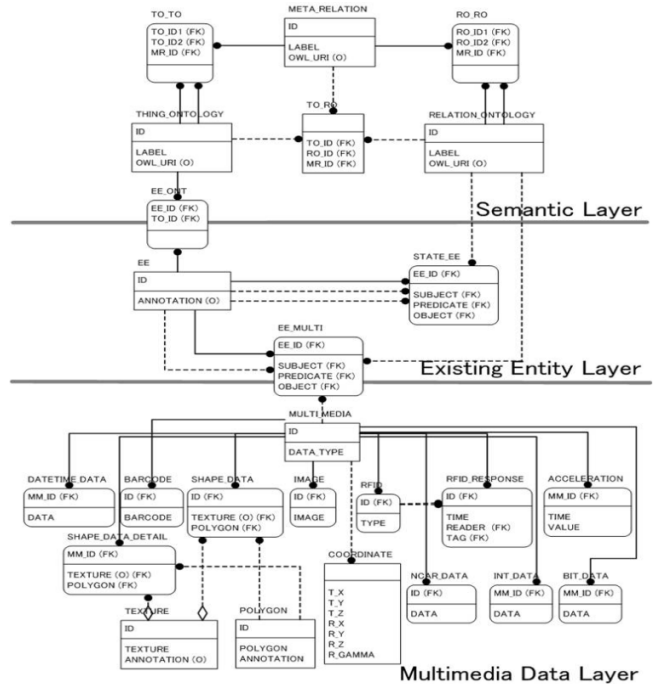


図1 CONSENT の論理スキーマ

ル THING ONTOLOGY に、プロパティがテーブル RELATION ONTOLOGY に当たる。意味情報が RDF スキーマを基礎として構築されることにより、CONSENT が扱える意味情報の範囲を RDF に帰属することができる。また、オントロジ記述言語として勧告された OWL[8] との親和性も高くなり、インポートやエクスポートが容易となる。なお、これらの意味情報は一般的な概念である「Upper Ontology」と空間固有である「Domain-Specific Ontology」で構成される。以下、THING ONTOLOGY, RELATION ONTOLOGY について詳述する。

● THING ONTOLOGY

本研究では実空間を対象とした意味情報を記述するので、THING ONTOLOGY はオブジェクトや人間の操作などの定義を蓄積する。THING ONTOLOGY 間の関係は RELATION ONTOLOGY に蓄積する。

具体的なインスタンスとしては、オブジェクト(「机」・「書類」・「PC」 etc.) や操作(「置く」・「書く」 etc.) といったものを固有の ID と LABEL, OWL での URI を属性として蓄積している。また THING ONTOLOGY は階層構造となっており、その階層は人間の概念に基づいた分類がされているので、「概念階層」であるということが出来る。この THING ONTOLOGY では、一般的な物体に対する定義やその IS-A 関係などが Upper Ontology で、さまざまなソースで公開されているものを利用することができる。さらにその下に空間固有である Domain-Specific Ontology を追加していくことで構成されている。

● RELATION ONTOLOGY

RELATION ONTOLOGY は RDF でのプロパティに相当する。RELATION ONTOLOGY は THING ONTOLOGY 間の関係を記述する。具体的には「本は机で読むことができる」・「この書類は個人机には置いてはいけない」などの表現を RELATION ONTOLOGY に蓄積する。このような関係を本論文では「WHAT」「WHERE」「ACTION」と定義する。例えば「本は机で読むことができる」では本が WHAT, 机が WHERE, 読むが ACTION である。

RELATION ONTOLOGY では THING ONTOLOGY に比べて、Domain-Specific Ontology が占める割合が多いため、構築の手間が大きい。

2.1.2 マルチメディアデータ

オブジェクトや操作に付随するマルチメディアデータは意味情報と独立して蓄積される。CONSENT ではすべての

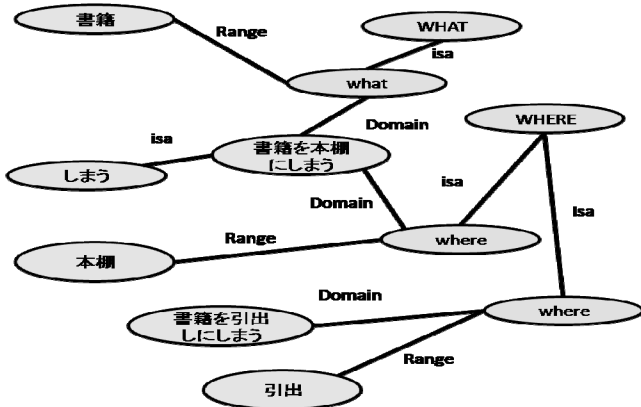


図 2 CONSENT でのオントロジ表現

マルチメディアデータに対して一意な ID を与えて管理する。これにより、マルチメディアデータが持つ意味を隠蔽し、オブジェクトや操作の意味情報との独立性を保証する。

オブジェクトに付随するマルチメディアデータとして、オブジェクトの画像や、オブジェクトに添付された RFID の情報、商品 ID などを蓄積する。操作に付随するマルチメディアデータとしては、RFID の反応や、加速度センサの値などを定義している。

2.1.3 存在エンティティ

上述の意味情報とマルチメディアデータの間には、多対多の関係が存在する。例として、同じ形でタイプ名称が異なるモノもあれば、同じタイプ名称で異なる形のモノがあるというのが一般的である。そこで、まず現実世界に実際に存在する「モノ」や「ユーザの行った操作」の一つ一つをエンティティとして捉え、「存在エンティティ」と呼び、それをアトミックなデータとして DB に蓄積する。意味情報とマルチメディアデータは独立のものと考え、両者の関係が実在することを存在エンティティによって表す。すなわち、存在エンティティは、ある意味情報(タイプ)に対して実在するインスタンスである。

例えば「Aさんの本」は、本という意味情報を持ち、バーコードや RFID タグ、一般的な書籍の形状といったマルチメディアデータを持っている、というオブジェクト「Aさんの本」が実在することを存在エンティティで表現している。CONSENT では、実在のオブジェクトに RFID タグや RFID リーダ、人に加速度センサを取り付け、それらの ID やセンサ値をマルチメディアデータとして登録し、DB内の存在エンティティと関連付ける。これにより、現実からのリアルタイムな状態取得を実現できる。

2.2. 概念共有環境 CONSENT の構築コスト

CONSENT における構築コストは前述した“意味情報”、“マルチメディアデータ”、“存在エンティティ”のインスタンスを取得・蓄積するコストだと言える。そこで、本節では本研究の対象となる CONSENT に関する意味情報を構築する上で発生するコストを挙げ、考察を行う。またここでいうコストとは構築をおこなう際にオントロジ編集者にかかる「手間」として定義する。CONSENT では意味情報の整合性や妥当性を維持するため、オントロジを記述することができるのはオントロジ編集者というエキスパートのみに限定する。

2.2.1 THING ONTOLOGY

THING ONTOLOGY に関しては、Upper Ontology は一般的なオブジェクトや操作なので、Web などでさまざまなソースが公開されている[9][10]。そこで本研究では JICFS[11]などをインポートし、空間用にカスタムすることで、容易に構築した。一方で Domain-Specific Ontology は、空間固有のオブジェクトや操作であるが、普段の生活の多くは一般的な行動なので、必要となった際にオントロジ編集者が追加することで解決することができ、頻度もあまり高くないのでコストはあまり大きくない。エンドユーザが新たなオントロジを必要とする場合は、オントロジ編集者に依頼することでオントロジの整合性や妥当性を確保する。

2.2.2 RELATION ONTOLOGY

RELATION ONTOLOGY は THING ONTOLOGY とは違い、Upper Ontology として利用できるものは少なく、ほとんどが Domain-Specific Ontology である。たとえば「本を机で読む」という行動も、空間によっては禁止されているかもしれない。このように RELATION ONTOLOGY は空間固有の知識やルール、暗黙の了解などを顕著に表現するので、その空間を熟知したオントロジ編集者への負担がおのずと大きくなっていく。

このような Domain-Specific Ontology を構築する際は、やはり手動で作成する割合が非常に多くなる。たとえば RDF のようなトリプルを記述することはクラスの記述、プロパティによる関連付けなど手間は非常に大きくなる。このような問題はオントロジ構築の大きな問題であり、普及を妨げる大きな要因となっている。

そこで本研究でも RELATION ONTOLOGY の構築コスト削減を重視する。

2.3 オントロジ構築支援

以上のことから本論文では、Domain-Specific Ontology の構築の手間を削減することで、オントロジ編集者の負担を軽減するとともに多くの有意なオントロジの作成を支援することを目標とする。そのための具体的な手法としてユビキタスドメインから得られる大量データを利用して、オントロジを半自動生成することでその目標を達成できると考えた。

ユーザの行動によるセンサデータと、一般的な Upper Ontology、さらにそれまでに蓄積された Domain-Specific Ontology の情報を組み合わせ、あらたな Domain-Specific Ontology を簡単に作成することができれば、空間のさまざまな行動をオントロジで表現可能となり、多くの支援を受けられ、支援を受けたユーザの行動のセンサデータからさらに新たなオントロジが作成される、というサイクルが生まれ、よりよいユビキタスドメインが構築できる。

そのために本論文では、Domain-Specific Ontology の割合が大きい RELATION ONTOLOGY の構築の手間を、ユーザの操作から得られたセンサデータを利用して削減する手法を提案する。これによって Domain-Specific Ontology 構築の手間の削減ができることを示す。

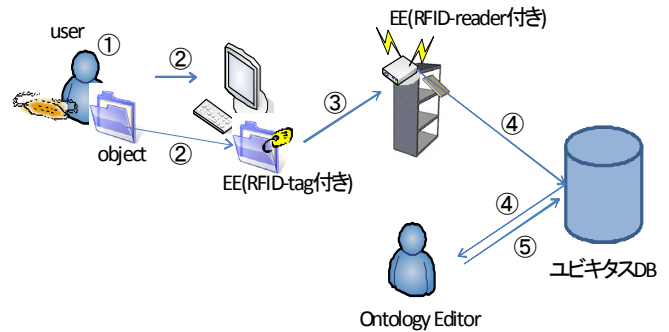
3. オントロジ構築支援フレームワーク

本章では、ユビキタスドメインにおけるオントロジ構築支援フレームワークを、CONSENT をモデルケースとして利用して提案する。なおここでは、Upper Ontology はアウトソースからすでにインポート・カスタムされているものとし、実運用でのセンサデータ (RFID, 加速度) を用いて Domain-Specific Ontology を動的に構築していく過程をサポートするフレームワークである。

3.1. フレームワーク

本論文で提案するフレームワークを図3に示す。まず環境には RFID や加速度センサが配置されていて、それらが実際のオブジェクトやユーザに結び付けられている。さらに Upper Ontology を用いた意味情報の紐づけを行っているため、最低限の操作の取得は可能であり、ユーザは普段通りの行動を行っている。

最初にユーザが新規オブジェクトを空間に持ち込む。そ



- ①Userがオブジェクトを空間に持ち込む
- ②UserがオブジェクトをDBに登録(形状と意味情報を結びつけ、EEとなる)
- ③Userはいつもの生活を送る(RFIDと加速度センサからなる操作履歴が蓄積)
- ④操作履歴から新たなオントロジ候補をOntology Editorへ推奨
- ⑤Ontology Editorは推奨結果を基に新たなオントロジをDBに登録

図3 オントロジ構築支援フレームワーク

してユーザはその物体を管理するために、意味情報と形状データや RFID タグといったマルチメディアデータを結びつけ存在エンティティ (EE) として DB に登録する。そうすると新たな存在エンティティが生成され、それをを用いた操作の取得が可能となる。

それらの操作の中には RELATION ONTOLOGY に関係が記述してないものも存在する。しかし、ユーザの操作から RFID によって「本と机」には何らかの関係があるらしいということが導出できる。たとえばユーザが実際に「本を机で読む」という操作を行ったとしよう。そのときこのオントロジが RELATION ONTOLOGY に記述してない場合、その操作の概念が空間にないという状態を表しているため、この操作は DB に蓄積できない。しかし「本」を表す RFID タグと「机」を表す RFID リーダは反応し、その反応の履歴は蓄積することができる。この履歴を用いることによって関係があると考えられる「対象」が取得できる。これは「WHAT」、「WHERE」、「ACTION」のうちの「WHAT」と「WHERE」を決定することができたことを表す。

さらに取得した対象間の関係を概念階層や操作履歴、加速度センサの値などから推奨する。つまり、ある「WHAT」と「WHERE」の組についてその間に存在すべき「ACTION」を推奨する。なおこの推奨には大きく分けて二つの異なる特徴を持つ推奨手法を提案する。

最後に推奨された結果をオントロジ編集者に提示し、推奨結果を編集・登録することで新たな Domain-Specific Ontology を構築することができる。最後にオントロジ編集者による承認・編集のフェーズを設けることで、オントロジの整合性・妥当性を維持する[12]。

3.2. オブジェクト登録

前述したように、CONSENT では意味情報とマルチメディアデータを存在エンティティを介して結びつけることでオブジェクトを管理している。よって持ち込んだオブジェクトに関する意味情報とそれに該当する形状データやそのオブジェクトに付加する RFID タグなどを登録する必要がある。この登録コストについては先行研究[13]で検討しているため、本論文ではその手法を用いてオブジェクト登録を行う。ここで、この新規オブジェクトに関する RELATION ONTOLOGY をユーザに提示することで空間知識やルール

などの共有が行えると考えた。そこで新規オブジェクトに関する RELATION ONTOLOGY と、そのインスタンスである操作が実際に全ユーザによって行われた回数などを提示し、ユーザの支援とする。例えば、新規オブジェクトとして「雑誌」が登録された場合、「雑誌を机で読む」や「雑誌を本棚にしまう」といった RELATION ONTOLOGY を全ユーザの操作回数とともに表示する。

3.3. 対象取得

次に新たな RELATION ONTOLOGY の候補である「対象」を取得するフェーズについて説明する。CONSENT ではオブジェクトに RFID タグ、特定の場所などには RFID リーダを設置している。これによって、「WHAT」と「WHERE」を決定し、RELATION ONTOLOGY のエントリを検索して、「ACTION」の候補をユーザに提示し、ユーザが実際に行った操作を取得している。またユーザに加速度センサを付け、C4.5 決定木などの分類木によって一意な「ACTION」を取得することでよりユーザの負荷を軽減するという研究も行っている[14]。しかし RELATION ONTOLOGY は Domain-Specific Ontology の占める割合が多く、「WHAT」と「WHERE」が決定してもそれに関するエントリが RELATION ONTOLOGY に記述していない場合がある。その場合はユーザの操作を取得できない。

そこで、RFID から取得した「WHAT」と「WHERE」の組を「対象」としてデータベースに蓄積し、それに関する RELATION ONTOLOGY を新規構築する。この対象はユーザの実際の操作から取得できるものなので信頼性が高い。またその際に加速度センサの値も蓄積し、構築に利用する。図4の例では「専門誌、共有テーブル」、「専門誌、机」に関する RELATION ONTOLOGY が存在しないので action が null となっている。そこでこの2組を対象として取得し、推奨を行う。

3.4. 対象間の推奨手法

本節では、取得された対象に関する RELATION ONTOLOGY を推奨するフェーズについて説明する。基本的な考え方としては、THING ONTOLOGY は概念階層なので、階層的に近いノード（兄弟ノードや親の親が同じノードなど）は、類似した概念を持っているということである。たとえば図5のような THING ONTOLOGY があつた場合、「SF」を基準ノードとした時に、親ノードが同じ「ファンタジー」「ホラー」などを「Level1 の類似オントロジ」と定義する。同様に親の親ノード（「書籍」）が同じ「スポーツ」などを「Level2 の類似オントロジ」と定義する。以下、階層をたどるごとに同様に定義する。すなわち Level が L の類似オントロジとは、基準ノードよりも L 階層上位のオントロジの子階層ノードすべてのうち、基準ノードを含む

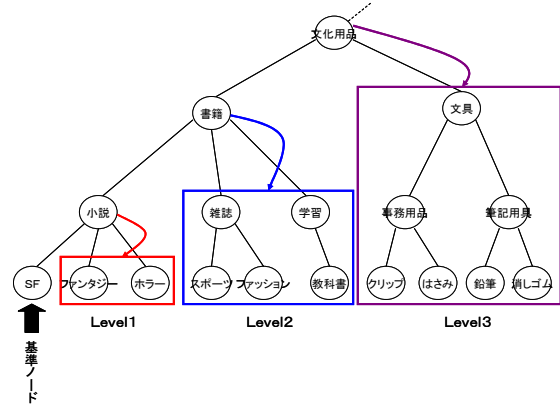


図5 概念階層と類似オントロジ

Level(L-1)のノードを除いたすべてのノードである。

そこで「WHAT の類似オントロジ」と「WHERE」、「WHAT」と「WHERE の類似オントロジ」に関する RELATION ONTOLOGY を参考に、「WHAT」と「WHERE」に関する RELATION ONTOLOGY をスコア付きで推奨する。スコアの計算方法については後述する。本論文では、RBO 法と RBH 法の二つの異なる特徴をもつ推奨手法を提案する。

3.4.1 RBO 法

RBO 法とは Recommend By Ontology の略で、オントロジ編集者が過去に記述したオントロジを用いて推奨を行う。

RBO 法の特徴は以下の通りである。

- 登録されたデータの構造に注目
- トップダウン

THING_ONTOLOGY のオブジェクトに関する要素集合を O、操作に関する要素集合を A とする。また概念階層において基準となるノードを x とした際に、Level が L の類似オントロジの要素集合を N_{Lx} と定義する。また WHAT が a、WHERE が b に関する RELATION ONTOLOGY を $R(a,b)$ とし、そのなかで ACTION が A_i であるものを $R(a,b)[A_i]$ とする。

$$N_{Lx} = \{O_1, O_2, \dots, O_n\} - (1)$$

$$R(a,b) = \{(O_1, O_2, A_1), \dots, (O_n, O_l, A_m)\} - (2)$$

するとある対象(a,b)に関する推奨のスコアは各 A_i ごとに以下ようになる。

$$S(L)[A_i] = \left(\frac{|R(N_{Lx}, b)[A_i]|}{|R(N_{Lx}, b)|} + W_{where} \times \frac{|R(a, N_{Lx})[A_i]|}{|R(a, N_{Lx})|} \right) \times W_{level(L)} + S(L-1)[A_i] - (3)$$

ここで $S(L)[A_i]$ はスコアであり、L は Level、 W_{where} は WHERE の類似ノードは WHAT に比べ類似度が低いという経験的な考えに基づく調整の重み、 $W_{level(L)}$ は Level が上がるごとに類似度は下がっていくと考え、その調整のための重み(距離関数)である。この二つの重みに関してはのちほどその妥当性を検証する。このスコアが式(4)で計算する閾値を超えている A_i のみが推奨される。

$$T_L = W_{level}^{(L-1)} \times (1 + W_{where}) \times t + T_{L-1} - (4)$$

what	where	action	who	MM_ID
学会誌	机	読む	A	null
重要書類	管理棚	しまう	B	1234
専門誌	共有テーブル	null	A	null
専門誌	机	null	B	1358
重要書類	共有テーブル	置く	C	1687

図4 操作履歴と対象

表 2 RELATION ONTOLOGY の構築コストの比較

	支援なし	支援あり
対象 (WHAT,WHERE) の組の選択	THING ONTOLOGY の組から手動選択	履歴から自動抽出
対象間の関係の選択	THING ONTOLOGY から手動取得	システムによって推奨

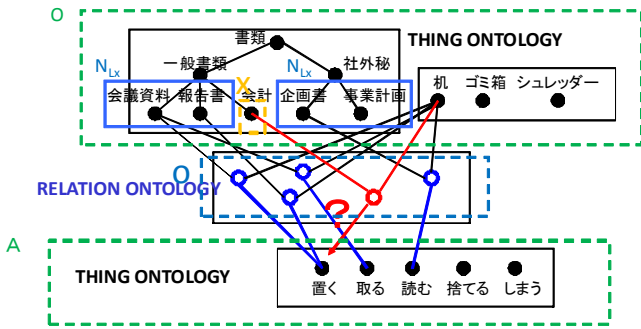


図 6 RBO 法の例

t はオントロジ編集者が設定できるパラメタで、「全ノード中の何割に記述されているノードを推奨させるか」を基本的な考えとして、設計する。

図 6 の例では、 x が「会計」となり N_{lx} が「会議資料」「報告書」となる。また $R(\text{会議資料}, \text{机})$ は「会議資料を机に置く」と「会議資料を机から取る」となる。すると対象(会計, 机)の「置く」に関するスコアは $L=1, W_{\text{level}}(1)=1, W_{\text{where}}=0$ とすると、 $S(1)[\text{置く}]=\frac{2}{2+0}=1$ となり、 $S(1)[\text{取る}]=\frac{1}{1+0}=0.5$ となる。ここで $t=0.6$ とすると $T_1=0.6$ となり、 $S(1)[\text{置く}] > T_1 > S(1)[\text{取る}]$ なので、「会計を机に置く」があらたな RELATION ONTOLOGY としてオントロジ編集者に推奨される。

3.4.2 RBH 法

RBH 法は Recommend By History の略で、ユーザが実際に行った操作の履歴を用いて推奨を行う。

RBH 法の特徴は以下のとおりである。

- 登録されたデータのインスタンスに注目
- ボトムアップ

$R(a,b)$ のインスタンス (実際の操作履歴) を $H(a,b)$ とし、 $R(a,b)[A_i]$ のインスタンスを $H(a,b)[A_i]$ とすると RBH 法でのスコアは以下ようになる。

$$S(L)[A_i] = \left(\frac{H(N_{La}, b)[A_i]}{|H(N_{La}, b)|} + W_{\text{where}} \times \frac{H(a, N_{Lb})[A_i]}{|H(a, N_{Lb})|} \right) \times W_{\text{level}}(L) + S(L-1)[A_i] - (5)$$

ここで RBH 法ではインスタンスをもつので、 $W_{\text{level}}(L)$ は κ 統計量[15]を用いて自動的に計算する。また対象(a,b)に関して、RFID とともに加速センサデータが取得出来ていた場合、そのデータを決定木にかけ、実際に行った操作を推定することができる。先行研究では、RELATION ONTOLOGY を用いることで推定の精度向上を図ったが、加速度センサデータのみからでも 70~80%の精度で推奨できるので[16]、ある程度は信頼できる推定結果が得られる。そこで加速度データが取得された対象に関して、そのデータから推定された A_i のスコアを加算する。

$$S(L)[A_i] = W_{\text{acc}} (S(L)[A_i]) - (6)$$

W_{acc} はどれだけスコアアップするかの関数である。これについても後ほど検証する。RBH 法はユーザの実際の操作を元に推奨ができるので、より空間の情報を反映したオントロジ構築ができる。

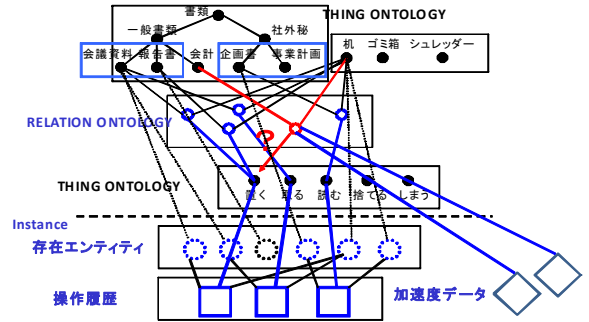


図 7 RBH 法の例

3.5 オントロジ構築の手間の比較

以上のようなフレームワークを用いた場合と、支援なしで RELATION ONTOLOGY の構築をする場合の手間を比較すると表 2 のようになる。

まず、対象を選択する際に、支援がない場合はすべてのオントロジの候補から手動で選択しなければならない、大きな手間となる。それに対してこのフレームワークでは、ユーザの行動から自動的にセンサデータが蓄積され、そのデータを用いて対象を自動取得することができるのでほとんど手間はかからない。

次に、対象間の関係の選択では、支援がない場合はすべてのオントロジから選択しなければならない。それに対してこのフレームワークを用いれば、ユーザの行動のセンサデータやそれまでに蓄積されたオントロジなどの情報を用いて推奨を行い、それを判断材料として選択をおこなうことで手間が削減できる。さらに、空間に蓄積された情報を基に推奨をおこなうので空間固有のオントロジの構築が可能である。

このように、このフレームワークを用いることで Domain-Specific Ontology の構築支援による手間の削減が可能となる。

4. 運用実験による評価

本章では、CONSENT を用いたフレームワークの評価を行う。

実験環境を以下に示す。

- 実験期間 2007/10/12-2008/1/20(継続中)
- 参加人数 17 人
- 実験場所 横浜国立大学 総合研究棟 E-306 室
- THING ONTOLOGY 4,631 個
 - Upper Ontology 3,965 個 (85.6%),
 - Domain-Specific Ontology 666 個 (14.3%)
- RELATION ONTOLOGY 4,885 個
 - Upper Ontology 1,100 個 (22.6%),
 - Domain-Specific Ontology 3,785 個 (77.4%)

- 存在エンティティ 265,196 個
- 履歴数 26,419 件

4.1. RBO 法の評価

RBO 法について、各種パラメタに対する推奨結果の評価を 10 分割交差検定を用い、Recall/Precision 及び f 値で行う。評価を行うパラメタは以下のように設定した。なおパラメタは式(3),(4)に準じる。

- 対象数 1514 組
- Level 1/2/3
- $W_{level}(L) = k^{(Level-1)}$ or $k * Level$
k: 0.1/0.2/0.5/0.7/0.8/0.9/1
- W_{where} 0.1/0.2/0.3/0.5/0.75/1
- t 0/0.1/0.2/0.3/0.5/0.75/1

まず $W_{level}(L)$ の妥当性の検証を行った。距離関数として、上の 2 つを用意してその影響を調べた。結果を図 8,9 に示す。このように $W_{level}(L)$ は k の値、および 2 つの距離関数に関わらず、スコアに大きな影響を及ぼさないことがわかった。よって式(3)は以下のように簡略化される。

$$S(L)[A] = \left(\frac{|R(N_{La}, b)[A]|}{|R(N_{La}, b)|} + W_{where} \times \frac{|R(a, N_{Lb})[A]|}{|R(a, N_{Lb})|} \right) + S(L-1)[A] - (3')$$

次に W_{where} の妥当性の検証を行った。結果を図 10 に示す。こちらもほとんど影響がないことがわかった。よって最終的に式(3)は以下のようになる。

$$S(L)[A] = \left(\frac{|R(N_{La}, b)[A]|}{|R(N_{La}, b)|} + \frac{|R(a, N_{Lb})[A]|}{|R(a, N_{Lb})|} \right) + S(L-1)[A] - (3'')$$

この 2 つの重み ($W_{level}(L), W_{where}$) の影響が現れなかった理由としては、今回の評価で使用したオントロジの構造によるものだと考えられる。今回のオントロジは階層ごとによる特異性があまり表現されておらず、同じような概念ばかりだったので、同様の推奨結果となり、重みが意味を持たなかったと考えられる。そこでさらなる運用を行って、空間独自のオントロジへと拡張していくことで、重みによる影響が出てくると考えられる。

この式(3'')に対して Level に対する評価を行った。結果を図 11 に示す。これにより Level が大きく推奨精度に影響を与えていることがわかる。推奨をしない場合に比べて、Recall は下がるが、Precision が大きく向上している。よって推奨なしに比べて、オントロジ編集者の手間を削減することができたとと言える。

4.2. RBH 法の評価

RBH 法についても、RBO 法と同様の手法で評価を行う。評価を行うパラメタは以下のように設定した。なおパラメタは式(5),(6)に準じる。

- 使用履歴数 26419 件
- Level 1/2/3
- W_{where} 0.1/0.2/0.3/0.5/0.75/1
- W_{acc} 加速度センサデータで推定された操作のスコアアップ関数
 - どのようにスコアアップするか
 $W_{acc}(x) = (1+k)(x)$, $W_{acc}(x) = (x+max)$
max: 推奨された最高のスコア
k: 0/0.2/0.5/0.8/1
- t 0/0.1/0.2/0.3/0.5/0.75/1

結果を図 12 に示す。このように RBH 法では RBO 法に

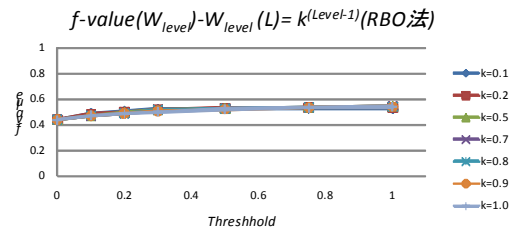


図 8 $f\text{-value}(W_{level}) - W_{level}(L) = k^{(Level-1)}$ (RBO 法)

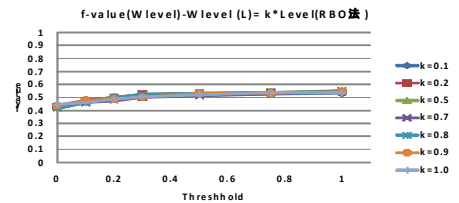


図 9 $f\text{-value}(W_{level}) - W_{level}(L) = k * Level$ (RBO 法)

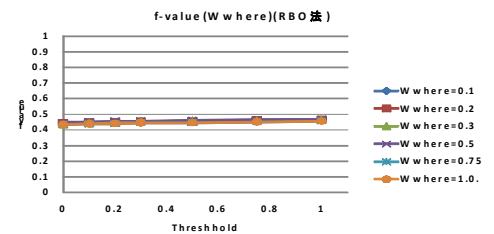


図 10 $f\text{-value}(W_{where})$ (RBO 法)

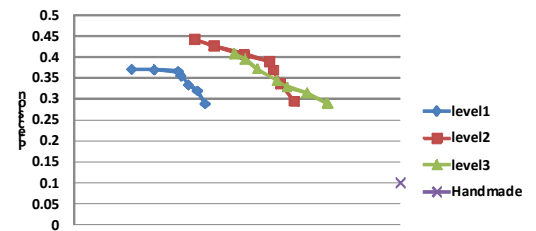


図 11 Recall/Precision(Level) (RBO 法)

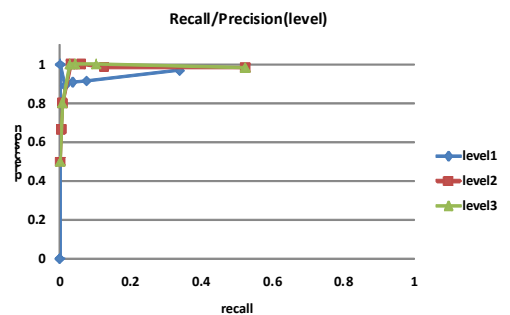


図 12 Recall/Precision(Level) (RBH 法)

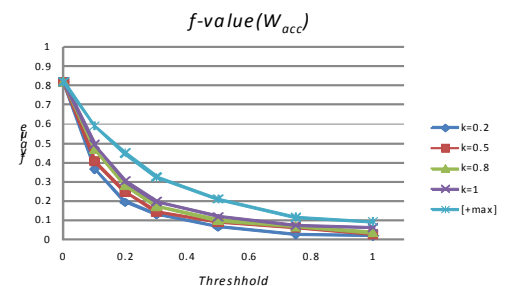


図 13 $f\text{-value}(W_{acc})$ (RBH 法)

比べて Recall は下がるが Precision は大きく向上している。これは RBH 法が実際にユーザが行った操作を基に推奨するため、正解が推奨される割合が RBO 法に比べて向上したと考えられる。これらの結果からオントロジ編集者に対してこの2つの推奨手法 (RBO 法と RBH 法) を提示することで、互いに Recall と Precision の値を補完しあうことが可能であることを示した。

また W_{acc} についても実験を行った。その結果を図 13 に示す。このように加速度センサデータを用いることでさらなる精度向上が実現できることがわかった。さらにここでは $W_{acc}(x) = (x+max)$ の関数を用いた場合が最も高い f 値となった。このことからより高いスコアを加算することで精度が向上することがわかった。これは加速度センサデータのみでの操作推定でも高い正解率が得られること、また候補となる操作自体が少なかったことが理由として考えられる。

4.3 作業時間の比較

最後にオントロジ構築支援を行う際に、このフレームワークを用いた場合と、支援なしで行った場合の作業時間を比較した。

50 組の RELATION ONTOLOGY を用意し、支援がある場合とない場合に対して、1 組を作成するために要した時間とその際の候補数を比較した。なおこの実験は2人の異なるオントロジ編集者が行い、ある人に特化したシステムではないことも示した。結果を表 3 に示す。このように人によって差異はあるものの、支援を受けない場合に比べて、短い時間で多くのオントロジの作成が可能になることを示した。これによって、単純な作業時間のレベルでも、手間の削減が可能になることを示した。

5. まとめと考察

本論文では、ユビキタス環境におけるオントロジ構築支援フレームワークを提案、実装し、その評価を行った。その結果、ユーザに負担をかけずに有用な情報を取得し、オントロジ編集者にあらたなオントロジ要素を提示することで、オントロジ編集者の手間を削減するとともに、空間の状態を反映したオントロジの作成が可能となることを示した。また推奨手法に関しても2つの異なる特徴を持つ RBO 法と RBH 法を実装し、ケースによって使い分けることでオントロジ編集者の助けとなるとともに、ロングテイル問題に関してでも対応できることを可能とした。

謝辞 本研究は平成 19 年度・平成 20 年度横浜国立大学教育研究高度化経費および平成 20 年度横浜国立大学大学院環境情報研究院共同研究プロジェクト経費の助成を受けて行った。

表 3 作業時間の比較

オントロジ編集者	支援なし	支援あり
A	22.76s/個 (18min58sec/50 個)	3.66s/個 (1hour27min17sec/1430 個)
B	28.35s/個 (9min27sec/20 個)	6.26s/個 (5min13sec/50 個)

文 献

- [1] SAMUEL R. MADDEN, MICHAEL J. FRANKLIN, JOSEPH M. HELLERSTEIN, WEI HONG, "TinyDB: an acquisitional query processing system for sensor networks", TODS, Volume30, Issue 1 (March 2005), pp.122-173
- [2] Arnab Bhattacharya, Anand Meka, Ambuj Singh, "MIST: Distributed Indexing and Querying in Sensor Networks using Statistical Models", VLDB '07, pp.854-865, September 23-28, 2007, Vienna, Austria
- [3] Amol Deshpande, Carlos Guestrin, Samuel R. Madden, Joseph M. Hellerstein, Wei Hong, "Model-Driven Data Acquisition in Sensor Networks", Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004, pp.588-599
- [4] 神崎正英. セマンティック・ウェブのための RDF/OWL 入門. 森北出版株式会社, 2005.
- [5] E. M. Maximilien, "Semantic Web Services for Human Activities.", In Proceedings of IEEE Services Computing Conference (SCC), pages 11-18, Chicago, IL, Sept. 2006. IEEE Computer Society.
- [6] 富井尚志. マルチメディアデータベースに基づく高度コミュニティ空間の実現. 電子情報通信学会誌, Vol. 89, No. 6, pp. 511-517, June 2006.
- [7] 南博康, 賀来健一, 富井尚志. 意図共有を実現するオフィス仮想環境の設計. 電子情報通信学会技術研究報告, Vol. 104, No. 345, pp. 1-7, Oct 2004.
- [8] World Wide Web Consortium. World wide web consortium. <http://www.w3.org/>, W3C Recommendation 10 February 2004
- [9] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening wordnet with DOLCE. AI Magazine, 24(3):13-24, 2003.
- [10] I. Niles and A. Pease. Towards a standard upper ontology. In The 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, 2001
- [11] 財団法人 流通システム開発センター, <http://www.dsri.jp/company/jicfsifdb/top.htm>
- [12] 高梨 勝敏, 佐藤 俊也, 原島 一郎, "Web コンテンツからのオントロジーの再構成方法の提案と試作", 人工知能学会論文誌, Vol.20 No.6, pp.417-425, Nov, 2005
- [13] 渡邊優作, 佐々木貴司, 富井尚志. クエリフィードバックによる意味情報推奨機構を有する成長型ユビキタス環境データベースの実装と評価. 情報処理学会論文誌データベース, Vol.50, No. SIG20 (TOD36), pp. 1-13, Dec 2007.
- [14] 猿田芳郎, 富井尚志, "加速度センサと RFID を用いたユビキタス環境での利用者コンテキスト推定手法", 日本データベース学会 Letters, Vol.6, No.3, pp.13-16, 2007.12.
- [15] Fleiss.J.L, "Statistical Methods for Rates and Proportions", John Wiley&Sons(1973), 佐久間 昭訳, "係数データの統計学", 東京大学出版会, 1975.
- [16] C. Randell and H. Muller. "Context awareness by analysing accelerometer data." In B. MacIntyre and B.Iannucci, editors, The Fourth International Symposium on Wearable Computers, pp. 175-176. IEEE Press, 2000.