

オントロジーマッチングによる異なる仮想空間の 3次元データの意味的な統合の実現

南 博康[†] 岡田直也 原嶋 優作^{*} 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 工学部電子情報学科

[‡]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: { † d03hc049 , * b0044114 , ‡ tommy }@ynu.ac.jp, okada@arislabs.dnj.ynu.ac.jp

あらまし これまで、我々はオントロジーを用いて3次元仮想コミュニティ空間に存在する3次元データに対する意味情報や操作の意図を抽出し明示的、体系的に定義しておくことで、意味情報と形状データを分離してDBで管理するモデル化手法を示してきた。通常、オントロジーの持つ性質として共通性、合意性が考えられる。これと3次元空間を組み合わせることで、一つのオントロジーで統合された3次元仮想コミュニティを設計することができた。この場合、一つのコミュニティ空間ではその空間独自の観点、興味で生成されたオントロジーが生成される。しかし、そのオントロジーは異なるコミュニティ空間の意味や意図を考慮して構築されていないため、異なるコミュニティ空間の3次元データオブジェクトに対しては意味を伴った検索や操作が出来ない。そこで本論文では、オントロジーのマッチングを行うことで異なるコミュニティ空間の3次元データオブジェクトに対しても意味や意図を持った検索を行えることを示した。

キーワード オントロジーマッチング, 空間データベース, コミュニティ空間

Realization of Semantic Integration of the 3-D Data in Different 3D Virtual Communities with Ontology Matching

Hiroyasu MINAMI[†] Naoya OKADA Yusaku HARASHIMA^{*} Takashi TOMII[‡]

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

^{*} Division of Electrical and Computer Engineering, School of Engineering, Yokohama National University

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

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

E-mail: { † d03hc049 , * b0044114 , ‡ tommy }@ynu.ac.jp, okada@arislabs.dnj.ynu.ac.jp

Abstract By extracting the semantic information and the intention of the operation which exist in the three-dimensional virtual community space with using ontology and defining it specifically systematically, we had shown the modeling technique that it was managed with DB with separating semantic information and 3-D shape data. Ontology has generally commonness, agreement as a nature. By combining ontology and 3-dimensional space, the 3-dimensional virtual community integrated by one ontology has been designed. In this case, the ontology is designed with a original viewpoint and interest in the space is generated in one community space. But it is not built in consideration of the meaning or intention of different community space, the reference or operation accompanied by the meaning cannot be performed to the 3-dimensional data object of different space. Therefore, we show that inquiry which had a meaning and an intention also to the 3-dimensional data object of different virtual space can be performed with ontology matching in this paper.

Keyword Ontology Matching, Spatial Database, Community Space

1. はじめに

近年の高度なグラフィックコンピュータの登場と広域情報ネットワークをもとに3次元グラフィックスやVRを用いた仮想美術館などのサイバーコミュニティ空間が実現されている。これらの仮想

空間をDBMSで管理すれば、ユーザは空間内での検索や、更新、シミュレーションなどの多様な操作を行うことができる。

仮想空間やVRシステムとDBを連動させたシステムとしてVWDB[1]があげられる。また、現実世界の時間的、空間的な情報を

モデル化して DB で管理するシステムとして RWDB[2]がある。これらによって、VR などの仮想空間を DB の強力な管理機能により扱うことが可能となった。

このような仮想空間で空間中のオブジェクトに対して、何であるのかを表す意味（オブジェクトが「机」である etc）や行うことができる操作を表す意図（書く、しまう、置く etc）を基にした検索が可能になれば、更に利便性が増すことになる。そこで、我々は対象となる仮想空間のオブジェクトの意味や意図を抽出し、これらを複数のユーザが共有し、仮想空間の形状データと分離して DB で管理するモデル化手法を提案してきた[3]。このような「概念化の明示的な規約」はオントロジーと呼ばれる[4][5]。

我々は、オントロジーによって仮想空間中のオブジェクトに対する意味や意図の共有が行われた仮想空間を高度コミュニティ空間と定義した[6]。一般的に「コミュニティ」とは、ある意思統一のもとに共通意識を持った人達の集まりのことを指す。つまり高度コミュニティ空間とは、コミュニティ内の人たちの共通意識を明示化したオントロジーとそれによって意味や意図を付加されたオブジェクトが共有された3次元空間を指す。

ここで問題となるのは「オントロジーはそのコミュニティ内の共通意識の明示化」ということである。コミュニティ内の独自の視点、興味で生成されたオントロジーは有用であり必要性は十分にあると考えられる。しかし、別のコミュニティ空間のユーザとの意味や意図の共有を目的として構築されてはいないため、別のコミュニティ空間のオブジェクトに対してオントロジーを利用した意味や意図にもとづいた検索ができなくなる。

そこで、別のコミュニティとオブジェクトに対する意味や意図の共有を考えた場合、コミュニティ空間ごとに構築されたオントロジーのマッチング[7]~[12]や統合[13]を行うことが必要となる。マッチングは、オントロジー間の共通とされる概念に対してリンクを作ることであり、統合は複数（一般的には二つ）のオントロジーから新しい階層構造、意味、意図の定義を持つオントロジーを構築することである。オントロジーのマッチングを図ることにより異なるコミュニティ空間のオントロジーとオブジェクトに対して検索が可能となる。また、オントロジーのマッチングに加えて仮想空間特有の情報であるオブジェクトの存在を表す「存在エンティティ」やオブジェクトが持つ「3次元形状データ」のマッチング[14]を行うことにより、「意味が同じで形状データが類似したオブジェクト」などのマッチングが取れ、「同意概念で大きさが類似形状データを持つオブジェクトを探せ」といった検索が実現できる。

これらのマッチングを行うマッチャーは通常一つの手法では信頼できる結果を返さない[11]。よって、複数のマッチャーをそれぞれの確かさを考慮しながら用いて、最終的なマッチング結果を得る必要がある。

以上より本稿では、別のコミュニティ空間のオブジェクトと意味や意図の共有を実現するために、コミュニティ空間のあらゆる情報のマッチングを行い、更に複数のマッチング結果を統合し、信頼性が高い結果を得る手法を提案する。更に、別のコミュニティ空間のオブジェクトに対して意味や意図を持った検索例を示し、マッチングにより別のコミュニティ空間の利用者と意味や意図の共有が行えることを示す。また、実際のコミュニティ空間同士のマッチングの実験を行い、結果に対する考察を行う。

2. 3次元空間モデル化手法

本章では、我々が提案している3次元仮想空間モデル化手法につ

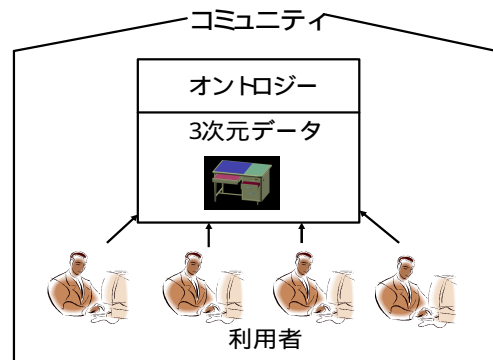


図1. オントロジーによるコミュニティ空間

いて説明する。一般にコミュニティとは、参加者がある程度共通の「意識」や「意図」を持った人の集まりである。例えば、ネット上に3次元仮想コミュニティを作成し、複数の利用者が参加する場合などを考える。

コミュニティでは、その背後にある「物の意味」、そこで行う「操作の意図」などが共有されていると考えられる。しかし、3次元仮想空間コミュニティにおいてオブジェクトに対する意味や操作の意図を扱おうと考えた場合、従来のVRMLなどの3次元空間モデル化手法では、意味や操作の意図はオブジェクトの形状データの追加属性値として表現されることで実現できる。

しかしこの方法では付加情報が多岐に渡り、かつ意味情報そのものが体系化されていないため以下の3点の問題が生じる。

- (1) 構造を持った意味の表現が困難
- (2) オブジェクトの付加属性内の文字列マッチングで意味や操作の意図が同じことを表現するが全検索となる
- (3) 異なるコミュニティ間で意味や操作の意図の情報の定義が違うことがある

この問題に対して、我々は「コミュニティ」に存在する意味や操作の意図を体系的かつ明示的に定義するオントロジーとオブジェクトの形状データを分離、対応させる概念設計手法を提案した[3]。(図1)具体的には、意味、操作の意図とオブジェクトの形状データを分離して管理し、その間をコミュニティ空間における存在を表す「存在エンティティ」で結びモデル化手法(3層構造モデル)である。以下それぞれの層について具体的に説明する。

2.1. 意味情報 (Ontology 層)

この層では、「コミュニティ」内に存在する意味や操作の意図から必要な概念を抽出し明示的かつ体系的にしておく。図2を用いて必要な概念を列挙すると

- ・ 「机」は「小物引き出し」と「書類引き出し」を部品として持つ。
- ・ 「小物引き出し」には「鉛筆」を中に「しまう」という機能がある
- ・ 「書類引き出し」には「参考書」を中に「しまう」という機能がある

このような意味や意図の構造をテキストの付加属性だけで実現す

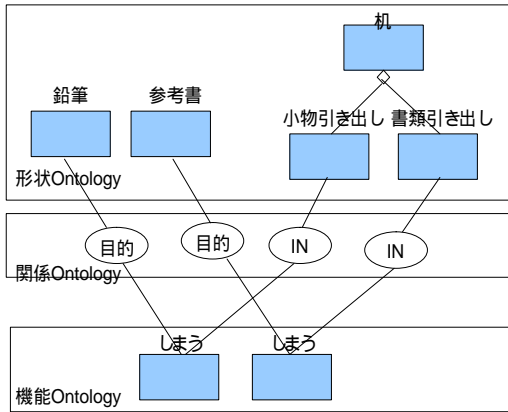


図2. 構造を持った意味情報

ることは困難である。したがって、コミュニティを構築する上で必要な意味や意図の情報を取り出し、体系的に明示的にしておく必要がある。このような「概念化の明示的な規約」は、オントロジー[4][5]と呼ばれる。

本モデル化手法では、オントロジーは(1)コミュニティ空間内で3次元形状の意味を表現する「形状オントロジー」と(2)コミュニティ内で実現される操作の意図を表した「機能オントロジー」(3)それらの関係を表した「関係オントロジー」(「人がモノを操作する」、「モノとモノとの位相関係」)の3種で表現される。以上をオントロジー層として蓄積していく。

2.2. 形状データ(raw data 層)

ここでは、コミュニティ空間を構築しているオブジェクトの形状データを蓄積しており、物体と物体の物理的な親子関係やポリゴンデータ(頂点データ、面ループ、色など)は全てこのraw data層で管理される。そのためraw data層に蓄積されるデータは、単なる形状データだけとする。つまり、どのポリゴン群が「机の引き出し」を表しているかのような意味、ポリゴン群に対する「机の引き出しに鉛筆をしまろ」などの操作は一切rawdata層には含まれない。これによりオントロジー層で蓄積した意味や意図の情報と形状データを分離できる。

2.3. 存在エンティティ(EE層)

意味情報による空間的問い合わせを実現するために、オントロジーによって明示化された意味を形状データと結びつける必要がある。我々は、オントロジーとraw Dataを直接結びつけるのではなく、オントロジーと形状データとの間を対応付ける仲介者として、「存在エンティティ」(Existing Entity: 以下EE)を導入する事を提案した。オントロジーが「意味の存在」を、raw dataが「形状データの存在」をそれぞれ示す一方で、存在エンティティは「仮想世界に個別に存在する」ことを示すエンティティである。このような要素を導入する利点は、文献[4]で述べられている。EE層もOntology層と同様に形状EEと機能EEから構成されている。形状EEは、3次元仮想空間中のオブジェクトの存在を指し、機能EEは実際に操作が行われている状態を表現する。

我々が提案している3層構造モデルでは、あるEEはあるOntology層のエンティティに属し、そのことは両層をまたぐリンクで表現されている。

以上から、図3に示すようにオントロジーと形状データをEEに

よって仲介することで、空間・形状データと知識情報の統合がなされる。

2.4. 意図付操作モデル

この節では、コミュニティのユーザが行う意図を持った操作を扱うモデル[15]を説明する。

これまで述べたモデル化手法により、操作の意図(知識)は機能オントロジーとして、操作の個別の状態は機能EEとして表現される。したがって、本コミュニティでの操作は(1)機能EE及び、(2)機能EEに結びつく形状EE及び機能オントロジーへのパス(図4の太線部分)の存在によって表現される。これにより、「あるノートをしまっている引き出しはどれ」など、その操作の存在、及びそれに関係のあるモノに対して検索が可能となる。ここで、「机の上のものをその机の引き出しにしまろ」などのよう「しまろ」と操作の意図を明示的に宣言する操作方法を明示的操作と定義する。明示的操作のPrimitive Activityは次のように表現される。

What × Where × How

Whatは操作対象となる形状EEの集合をさす。Whereはそのモノと空間的に関係を持つオブジェクトの集合であり、操作対象となる形状EEと空間的に関係を持つ形状EEの集合をさす。Howはオブジェクトとオブジェクトの関係集合であり、機能EEをさす。機能EEは操作が行われて初めて存在する。したがって、操作とはHowの要素(機能EE)を新たに作成し、関連するWhom、Where及び機能オントロジーにパスをはる、つまり図の太線部分を生成することにより実現される。具体的な操作の更新手順は文献[15]で述べられている。

3. 高度コミュニティ空間のマッチング

2章で述べた設計手法で構築したコミュニティ空間では、その空間独自の観点、興味でオントロジーが生成される。オントロジーは別のコミュニティ空間のオブジェクトの意味や意図を考慮して構築されていないため、別のコミュニティ空間のオブジェクトと意味や意図の共有ができない。しかし、高度コミュニティ空間のあらゆる情報(オントロジー、EE、形状データ)のマッチングを取ることができれば、別のコミュニティ空間へ意味や意図を持った検索が可能となり情報共有が実現できる(図5)。その際に、複数のマッチャーを用いてその結果を統合することで精度の高い情報共有結果を得ることを考える。以下本章では、3.1節でマッチングを行うマッチャー、3.2節で異なるコミュニティ空間同士の情報共有の実現についてそれぞれ説明する。

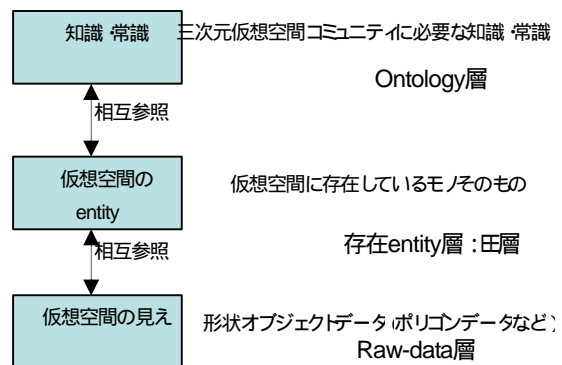


図3. 3次元仮想空間モデル化手法

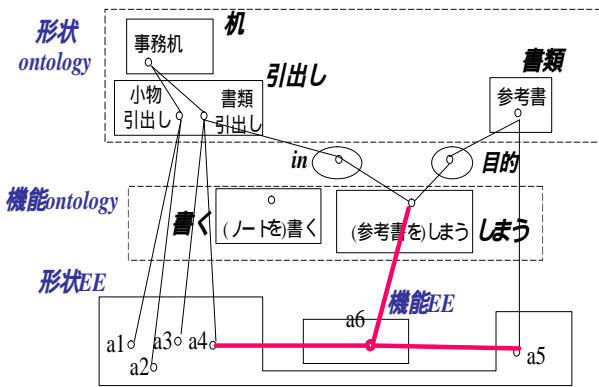


図4．操作概念

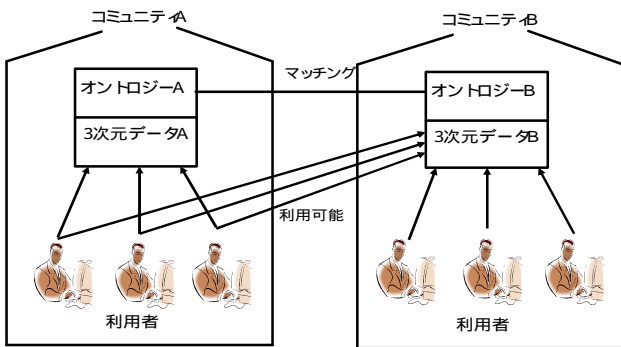


図5．マッチングによるコミュニティ空間の統合

3.1. マッチャーの種類

本節では、コミュニティ空間のそれぞれの情報（オントロジー、EE、形状データ）のマッチャーについて説明する。

オントロジーのマッチャーは、これまでタイプ（3次元仮想空間モデルのオントロジーのタイプ）[7]～[12]とインスタンス（3次元仮想空間モデルのEE）[7][10][12]の情報を利用したものが提案されている。これらは、マッチャーが類似すると判定したタイプの組み合わせを出力する。一般的に、「タイプの情報を利用するマッチャー」（以下：Type Based Matcher：TBM とする）は、タイプのラベルのスペルマッチ[7][8][11][12]、データ型やデータレンジ[9][12]、シソーラスに代表される外部知識の利用[8][9][10]などの制約を利用してマッチングを行う。「インスタンスの情報を利用するマッチャー」は、インスタンスに出現するキーワードの頻値、値の類似性（電話番号と年齢を表す数字データの類似性など）[7][12]を用いる。図6は、オントロジーのマッチャーによって同意概念とされたタイプが示された結果である。

本研究では、TBMに加えてEEの情報を利用するマッチャー（以下EBMとする）形状データを利用するマッチャー（以下SBMとする）を用い、同意概念の中から類似の状態や形状データを持つEEのみを選別する。つまり、EBMやSBMを利用することによって、EEから類似形状データを持つモノや類似の状態のEEを選択できるようになる。これによって、実際の世界におけるオブジェクトの形状や状態も踏まえたオントロジーマッチングを実現できると考えた。

EBMは、属性値に対する条件設定やある動作が行われている状態（機能EE）などを使用する。図7では、EEのa1とb1がそれぞれ(a1-a4-a3)、(b1-b4-b3)のパスによってモノがしまわれている共通の状態になっていることを示している。これによって、別のコミュニテ

ィ空間に対して共通の状態にある、もしくは過去に共通の状態であったオブジェクトの検索が実現できる。

SBMは、形状データの類似性を使用する。3次元オブジェクトの特徴量に関しては、[14]で提案されているものに基づいて形状データの類似性を判定すれば良い。図8では、同じ形状データを持つEEがマッチした例である。

以上の3種類のマッチャーを用い、違うコミュニティ空間に対して「同意概念で大きさが同じ形状データを持つEEを探せ」などの検索が可能となる情報共有を実現する。

3.2. 異なるコミュニティ空間の情報共有の実現

本節では、前節で説明したTBM、EBM、SBMを用いた別のコミュニティ空間との情報共有を説明する。

今回提案する手法では、TBM、EBM、SBMを用いることで、あるタイプと別のコミュニティ空間のEEとの間に新しいInstance of 関係を作成することとした。これによって、別のコミュニティ空間のEEを元のコミュニティ空間のタイプのインスタンスにできるため、情報共有が実現できる。以下、TBM、EBM、SBMそれぞれについて新しいInstance Of 関係の導出方法について説明する。

TBMではあるタイプに所属する全てのEEが別のコミュニティ空間の同意概念とされたタイプと新しいInstance of 関係が作られる（図9）。

EBM、SBMではあるタイプに所属して特定の条件を満たすEEが別のコミュニティ空間のタイプとInstance of 関係を作る（図10）。

これらの複数のマッチャーを用いて信頼度が高い新しいInstance of 関係を得ようとした時に、個々のマッチャーから得られるInstance of 関係の確かさはそれぞれ違う問題がある。したがって、

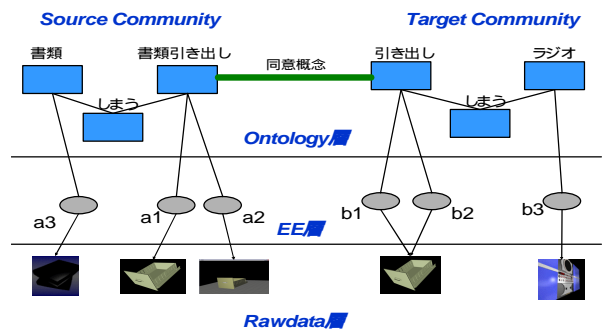


図6．TBMの結果

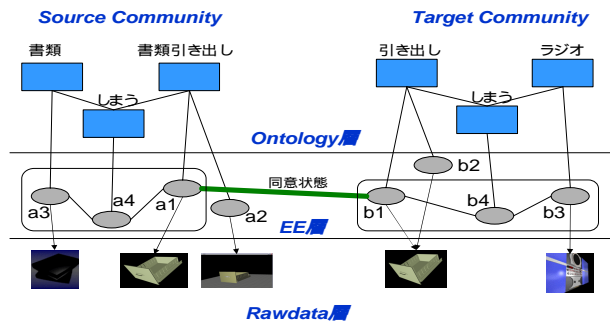


図7．EBMの結果

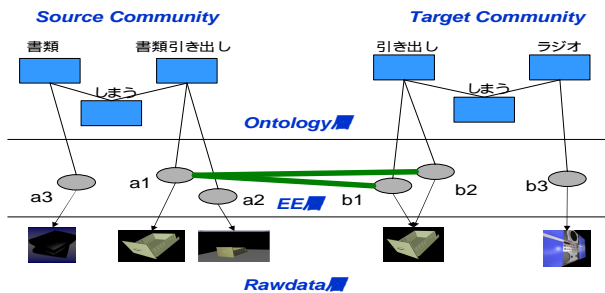


図 8 . SBM の結果

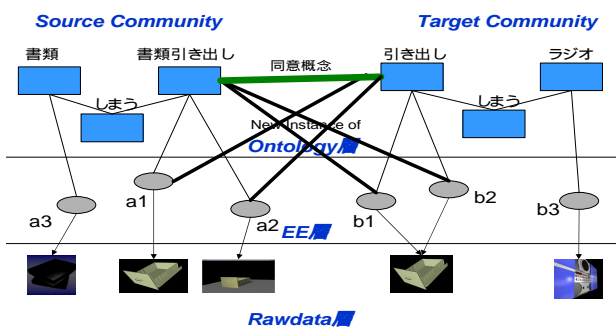


図 9 . 新しい Instance of 関係(TBM)

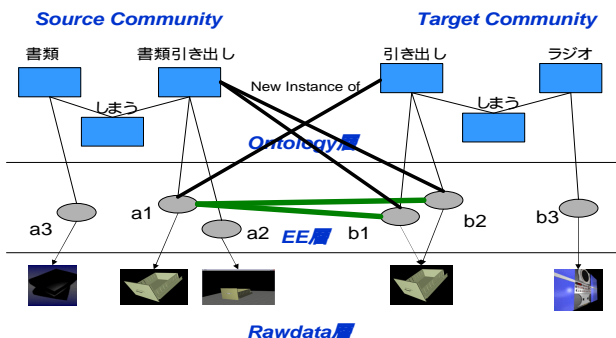


図 10 . 新しい Instance of 関係 (EBM, SBM)

マッチャーごとに得られる新しい Instance of 関係について、その関係の強さを複数のマッチャーの結果を統合し算出する。この強さは、「そのインスタンスがマッチングの結果、別のタイプのインスタンスとなり得る」ことを表す。これを正規化された1つの値である BOT(Belongs to Type)として表すこととする。ここで、 $0 \leq BOT_{it} \leq 1$ とする。 i はあるエンティティのIDを指し、 t は別コミュニティ空間のタイプのIDを指す。図11は、新しい Instance of 関係について BOT が導出された状態である。BOT に対してしきい値を設定すれば、適切な EE を別のコミュニティ空間から取得できることが可能となる。図11では、BOT を基に a1 が Target Community の「引き出し」として選択され利用可能となっていることを示している。BOT の導出方法については、次章で詳しく説明する。

4. 複数のコミュニティ空間との情報共有

本章では、(1) マッチングを利用した複数のコミュニティ空間との情報共有 (2) 情報共有に必要なデータの導出、(3) 新しい

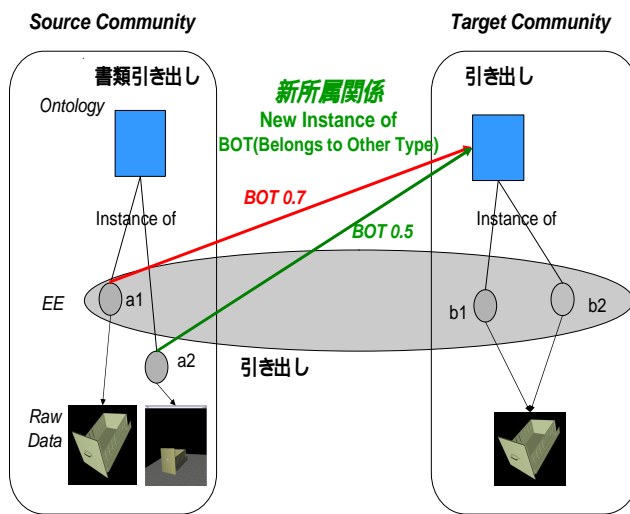


図 11 . 新所属関係の強さ (BOT)

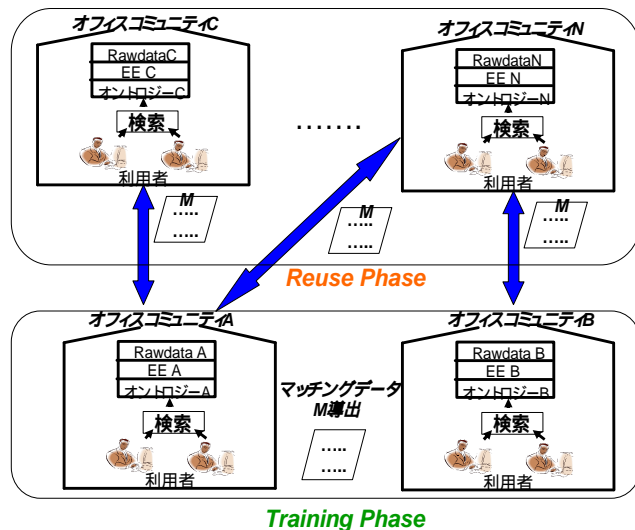


図 12 . 複数のコミュニティ空間との情報共有

Instance of 関係を利用して別のコミュニティ空間への具体的な検索方法を説明する。

以下、4.1節で複数のコミュニティ空間との情報共有、4.2節でマッチングに必要なデータの導出、4.3節で具体的なコミュニティ空間での検索事例について説明する。

4.1. 複数のコミュニティ空間との情報共有

本節では、複数のコミュニティ空間との情報共有について説明する。通常、コミュニティは複数存在する。例えば、大学などでは複数のゼミや研究室、会社などでは複数の課や支社がある。しかし、複数のコミュニティ空間ごとにマッチングを行っているにはコストが多くかかる。

よって、マッチングに必要なデータをトレーニングフェーズで導出して、それをリユーズフェーズで使いまわすことによりコストの削減を実現する(図12)。マッチングに必要なデータの導出については、次節で詳しく説明する。

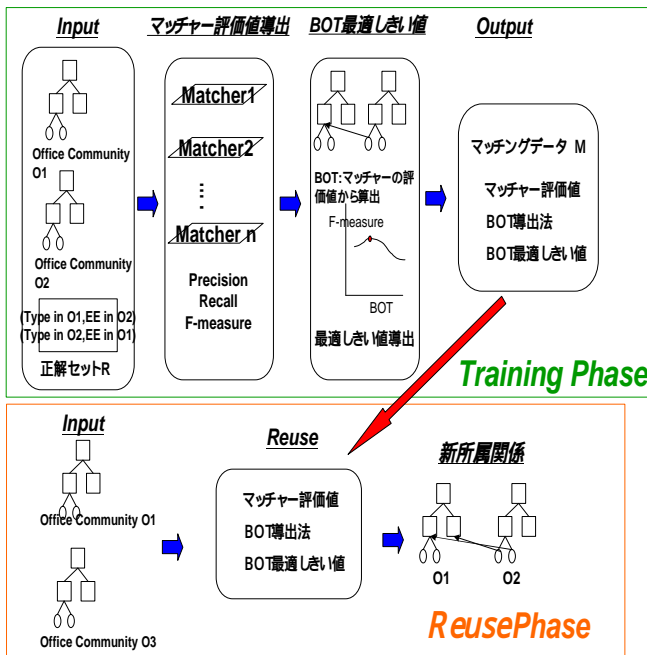


図 1.3 . トレーニングフェーズとリユーズフェーズ

4.2. マッチングに必要なデータの導出

本節では、トレーニングフェーズでマッチングに必要なデータの導出方法を説明する。マッチングに必要なデータは以下である。

- (1) マッチャーの信頼度
- (2) 複数のマッチャーの信頼度からの BOT の導出方法
- (3) BOT に対する最適しきい値の設定

まず(1)から説明する。図9と図10では、返す結果が違っているため、それぞれの結果を基にマッチャーの信頼度を設定する。信頼度については、適合率 (Precision) 再現率 (Recall) F-measure[16], Overall[17]などが考えられる。F-measure、Overall はそれぞれ以下の計算式で求めることができる。

$$F - measure = \frac{2 * precision * recall}{(precision + recall)}$$

$$Overall = recall * (2 - \frac{1}{precision})$$

これらを求めるために、各マッチャーの結果を手動で用意した正解セット R と比較し、信頼度を導出する。

次に(2)について説明する。マッチャーの信頼度を基に新しい Instance of 関係ごとに BOT を導出する。図9、10、11を用いて説明する。図9のマッチャーの信頼度が1、図10のマッチャーの信頼度が0.4であるとし、これらのマッチャーから得られた新しい Instance of 関係について BOT を算出する。BOT の算出方法については、マッチャーの信頼度の単純和、相加平均、相乗平均などが考えられる。BOT の導出方法について相加平均を用いた場合、新しい Instance of 関係 A1 (a1, 引き出し) A2 (a2, 引き出し) の BOT はそれぞれ次のようになる。

$$A1 = (1 + 0.4) / 2 = 0.7$$

$$A2 = (1) / 2 = 0.5$$

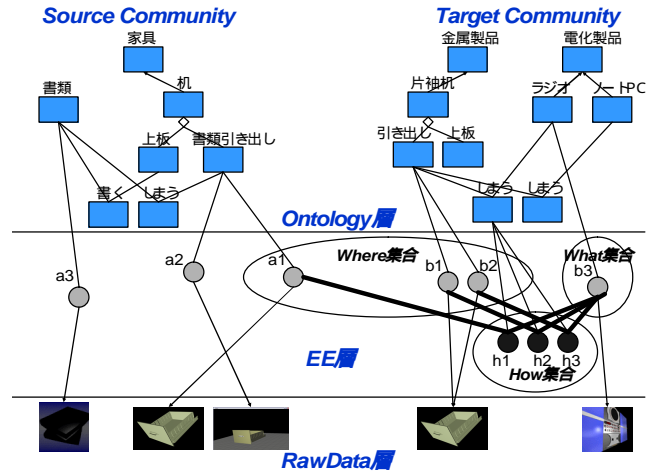


図 1.4 . 別のコミュニティ空間への検索例

最後に(3)について説明する。BOT が求めたそれぞれの新しい Instance of 関係について正解セット R と比較を行いながら BOT の最適しきい値を求める。図11では、最適しきい値を求めた結果、a1 が Target Community の「引き出し」として選択されている。

以上の3つのデータをトレーニングフェーズで導出し、リユーズフェーズで再利用する。トレーニングフェーズ、リユーズフェーズの関係を図13に表す。

4.3. マッチング結果を用いた具体的な検索例

この節では、マッチングの結果を用いてコミュニティ空間から別のコミュニティ空間の EE に対して、意味や意図を持った検索が拡張できることを具体的な検索例によって説明する。

図11の a1 が Target Community の「引き出し」として選択された状態から Target Community の利用者が以下の操作を表す検索を行った場合を考える。

検索：ラジオを全ての引き出しにしまえ (Source Community にも検索を拡張する)

検索は2.4節で述べた What、Where、How の集合を特定し、次の手順で行われる。

(1) Where、What の特定

図12から「引き出し」には、{b1,b2}と Source Community の{a1}が所属する。そのため、Where 集合は{a1,b1,b2}となる。What 集合は、ラジオの EE{b3}となる。

(2) しまわれた状態の作成

コミュニティ空間で行われている実際の操作は、機能オントロジーに所属する機能 EE を How 集合として作成し、What、Where 集合にパスを作ることで操作が行われた状態が作成される。今回の例では、How 集合として、{h1,h2,h3}を作成し、Where 集合{a1,b1,b2}と What 集合{b3}にパスを作ることで「引き出し」である{a1,b1,b2}にラジオ{b3}がしまわれた状態が作られる (図13)。

Where 集合だけでなく、What 集合についても上記の手順を踏めば、別のコミュニティ空間の EE を検索可能となる。つまり、Where、What 集合にマッチングの結果を基にして、別のコミュニティ空間の EE を要素として加えれば、別のコミュニティ空間の EE に対して意味や意図を持った検索を行うことが可能となる。複数のコミュニテ

イ空とのマッチングを取ることで、複数のコミュニティ空間から EE を取得することも可能となる。

5. マッチングの評価

本章では、実際のマッチングを行い、4.3 節で示した手法の有用性を検証する。

5.1 節で実際のコミュニティ空間トレーニングフェーズを行い、マッチャーの信頼度、BOT の最適しきい値の導出、5.2 節でリリースフェーズを行いマッチャーの信頼度、BOT が別のコミュニティ空間とのマッチングで利用できることを検証する。

5.1. マッチャーの信頼度、BOT の最適しきい値の導出

今回の実験に用いたのは4つのオフィスコミュニティ空間である。それぞれを以下に示す。

表1. コミュニティ空間

コミュニティ空間	Type数	EE数	Rawdata
A	19	155	3ds max
B	32	184	3ds max
C	25	31	3d model search engine
D	47	204	3d model search engine

A は、我々の研究室をメンバー合意のもとに作成した。B、C、D はそれぞれ異なるコミュニティ空間の管理者が作成した。オントロジーは、EDR 概念辞書[18]を基に作成し、足りない意味や操作の意図については管理者が付け加えた。rawdata は、A と B では 3ds max[19]を用いて作製、C と D では 3D Model Search Engine [20]から用意した。

マッチャーは、以下を用意した。

表2. TBM

マッチャー	補助情報
スペルマッチャー	-
パススペルマッチャー	-
カタログナンバーマッチャー	辞書の使用

表3. SBM

マッチャー	補助情報
面積マッチャー	-
体積マッチャー	スカラー三重積の利用

次に各マッチャーについて説明する。

・ スペルマッチャー

[7][9][11][12]で用いられている。タイプのラベルに対して、1-gram で解析しタイプ間の類似度を算出する。類似度は以下の計算式を用いる。

$(\text{タイプ間の共通文字数}) / (\text{二つのタイプ全文字数})$

「書類引き出し」と「引き出し」の類似度は、8/10 となる。

・ パススペルマッチャー

[11]で用いられている。タイプ間の類似度を、オントロジーのパスに対して 1-gram で解析し算出する。図13の「書類引き出し」は「家具 / 机 / 書類引き出し」、「引き出し」は「金属製品 / 片袖机 / 引き出し」のパスで表現される。これらのパスに対して「書類引き

出し」と「引き出し」の類似度を以下の計算式から算出する。

$(\text{パス間の共通文字数}) / (\text{二つのパス全文字数})$

「書類引き出し」と「引き出し」の類似度は、10/20 となる。スペルマッチャーに比べて階層構造の分類が類似度に反映される。

・ カタログナンバーマッチャー

[8][10][11]で用いられている。シソーラスなどの外部の知識体系を利用する方法である。今回は、日本標準分類[21]のナンバーをタイプに持たせてナンバーが一致するものを同意概念とした。

・ 面積マッチャー

[14]にある方法に基づき、ポリゴンの面積を求めた。あるタイプ t_1 に属する形状データの面積の平均、標準偏差を算出し、別空間の EE で以下の値に収まるものを t_1 に属することができるとした。

平均 ± 標準偏差

・ 体積マッチャー

[14]にある方法に基づき、ポリゴンの体積を求めた。面積と同様の方法を用いた。

まず、トレーニングフェーズをコミュニティ空間 A、B でを行いマッチングに必要なデータを導出する。手動で正解セットを用意しマッチングを行って、各マッチャーの信頼度と BOT の最適しきい値を算出する。スペルマッチャー、パススペルマッチャーについては、類似度に対するしきい値を変化させて F-measure を測定し、最も良い F-measure の値を信頼度とした。以下に A と B のマッチングで得られたマッチャーの信頼度を示す。

表4. マッチャーの信頼度

	類似度に対するしきい値	Precision	Recall	F-measure (信頼度)
スペルマッチャー	1	0.903	0.659	0.762
パススペルマッチャー	0.85	0.863	0.552	0.673
カタログマッチャー	-	0.341	0.524	0.413
体積マッチャー	-	0.777	0.671	0.720
面積マッチャー	-	0.612	0.851	0.712

スペルマッチャーが高い信頼度であるのは、オントロジー構築に EDR を使用し、タイプに共通のラベルが多く用いられたためである。カタログナンバーマッチャーは、日本標準分類の約 36000 種の分類の中からタイプに適したものを選択し一致する確率が低くなったため、信頼度が低くなった。面積と体積については、タイプのラベルが違うが、同じ形状や大きさの形状データを持つ EE を発見できていた。

この結果をもとにして、A と B で作られた新しい Instance of 関係

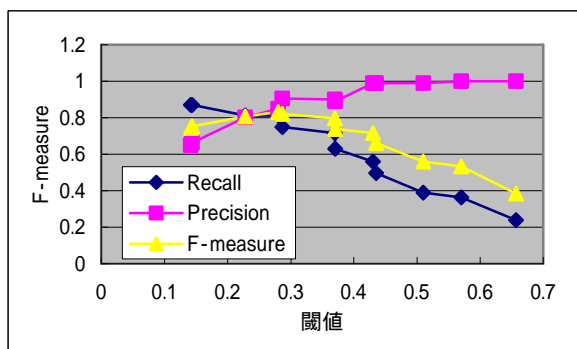


図15. BOT のしきい値との関係

について BOT をマッチャーの信頼度の平均から算出し、更に BOT の最適しきい値を求めた。図 15 に BOT のしきい値と Precision、Recall、F-measure の関係を示す BOT が 0.28 の時に F-measure が 0.831 と最大になった。

5.2. マッチャーの信頼度、BOT の利用

本節では、リユースフェーズを行う。リユースフェーズは、A と C、A と D で行い、表 4 のマッチャーとその信頼度を用い、BOT の最適しきい値は前節で求めた 0.28 とした。結果を以下に示す。

表 5 . A と C、A と D のマッチング結果

	Precision	Recall	F-measure
AとC	0.946	0.854	0.898
AとD	0.686	0.890	0.775

両方の結果とも得られた F-measure の値は高く、マッチャーの信頼度と BOT の最適しきい値を使い回すことが可能であることを示している。原因として、EDR 辞書をオントロジーの構築に利用しているために共通の概念(タイプ)をスペルマッチャーやパススペルマッチャーで発見しやすかったことが考えられる。A と C、A と D の結果の違いは、A と C では、C の EE が少なくマッチャーが間違いをあまり出力しなかったため Precision が高くなったこと、逆に A と D のマッチングは、D の EE が多くマッチャーが間違いを多く出力したために Precision が下がったことが原因であると考えられる。

正解セットで出力できなかったものは、「教科書」と「マニュアル」など形状は似ているが「タイプ」のラベルが違うものである。形状データの面積、体積だけでなく[14]で挙げられている他の特徴量もマッチャーとして使用すれば改善が期待できる。

間違えて出力したものは、「タイプ」のラベルが同じで形状が違うものである。改善のためには、スペルマッチャーなど「タイプ」のラベルを利用したものだけでなく制約条件や階層構造を利用したマッチャーを用意する必要があると考えられる。

6. まとめと今後の課題

本論文では、(1)一般的な TBM に加えて EBM、SBM を用いることによってインスタンスレベルでのマッチングの実現(2)複数のマッチャーの結果を用いてより適切な EE の取得(3)マッチング結果を用いた別のコミュニティ空間への検索方法(4)マッチャーの信頼度、BOT 導出法、BOT の最適しきい値の使い回しによるマッチングコストの削減が可能なることを実験により示した。

今後の課題としては、(1)新しいマッチャーの用意(2)マッチャーの評価方法や BOT の算出方法を変えた実験(3)マッチングが取れなかったオブジェクトに対する検索などが考えられる。

謝辞

本研究の一部は文部科学省科学研究費補助金(課題番号 40313473)による。

本研究の遂行にあたり EDR 電子化辞書「概念辞書」を利用した。

文 献

[1] 渡辺知恵美, 増永良文: "仮想世界データベースシステム VWDB2 における仮想世界同期法", 情報処理学会論文誌, Vol.44, No.SIG8(TOD18), pp.65-77, Jun. 2003

[2] T. Tomii, K. Salev, S. Imai, and H. Arisawa, "Human Modelling and Design of Spatio-Temporal Queries on 3D Video Database," Visual

Database Systems 4 (VDB4), L'Aquila, Italy, 27-29 May 1998.

[3] 岡田直也, 富井尚志, "ontology を用いた空間形状データの意味情報モデリング", データ工学ワークショップ論文集 (DEWS2003), Mar, 2003 .

[4] JOHNF.SOWA, "Knowledge Representation :Logical, Philosophical, and Computational Foundations", Brooks/Cole,2000.

[5] Gruber, Thomas R., "A translation approach to portable ontology specifications. In Knowledge Acquisition, vol.5, pp199-220, 1993.

[6] 富井尚志, "高度コミュニティ空間の構成方式", 情報処理学会研究報告, DBS-131(II), pp.33-40, Aug, 2003 .

[7] Doan, A.H., J.Madhavan, P.Domingos, A.Helevy, "A Learning to Map between Ontologies on the Semantic Web", WWW2002

[8] P.Mitra, G.Wiederhold, M.Kersten, "AGraph-Oriented Model for Articulation of Ontology Interdependencies," In Advances in Database Technology (EDBT2000), LNCS1777, pp.86-100, Springer Verlag, 2000.

[9] Madhavan J, Bernstein PA, Rahm E, "Generic schema matching with Cupid." In: Proc 27th Int Conf On Very Large Data Bases, pp. 49-58, 2001

[10] David W. Embley, David Jackman, Li Xu, "Multifaceted Exploitation of Metadata for Attribute Match Discovery in Information Integration", WIIW, 2001

[11] Hong-Hai Do, Erhard Raham, "COMA - A system for flexible combination of schema matching approaches", Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002

[12] Li W, Clifton C, Liu S, "SemInt: a tool for identifying attribute correspondences in heterogeneous databases using neural network", Data Knowl Eng 33(1):49-84, 2000

[13] Noy, N. F. and M. A. Musen, "PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment" In the Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000), Austin, TX. Available as SMI technical report SMI-2000-0831, 2000.

[14] 村上存, 湯浅修二, 田島史和, 斉藤祐樹 "形と色の特徴量を用いたデザイン支援に関する研究", 日本機械学会第 11 回設計工学・システム部門講演会講演論文集, pp167-171

[15] 高木良成, 岡田直也, 竹島広人, 谷岡岡, 富井尚志, "DB で管理された 3 次元仮想コミュニティにおける意図付操作モデルの導入" 信学技報, Vol. 103, No.356, DE2003-111, pp.1-6, 2003.10.

[16] C. J. "Keith" van Rijsbergen "Information Retrieval", <http://www.dcs.gla.ac.uk/Keith/Preface.html>, 1979

[17] H.DO, S.Melnik, E.Rahm, "Comparison of Schema Matching Evaluations", Workshop on Web Databases, 2002

[18] 通信総合研究所, EDR 概念辞書, <http://www.crl.go.jp/overview/index-J.html>

[19] 3ds max, <http://www.discreet.com/3dsmax/>

[20] 3D Model Search Engine, <http://shape.cs.princeton.edu/search.html>

[21] 総務省統計局基準部編, 日本標準分類, 財団法人全国統計協会連合, 1990