建築資料のデジタルアーカイブ化におけるオントロジーの構築とその役割

○ 研谷紀夫¹ 藤原正仁² 馬場章³

Construction and role of the ontology for creating digital archives of architectural material

○ Norio TOGIYA, Masahito FUJIHARA, Akira BABA

We constructed digital archive of architectural material of Seitaro Tsuboi house which was built in Tokyo 1923. However, the architectural material contains many technical terms and a large number of names of people and organizations. If we did not clarify the relationship among the original, the digital archive would not make it easy to understand the relationship among the material. As a result, we developed an ontology concerning the knowledge information described in the architectural documents, and created archive based on the ontology. Application of ontology is prominent in industrial, medical, and legal fields, but there are few cases where it has been applied to historical material. Consequently, this paper explains the merits of ontology developing in the digital archiving of historical material.

1. 概要

東京大学大学院情報学環・学際情報学府では、2004 年より東京都北区にある坪井誠太郎邸の建築資料のデジタルアーカイブ化を行うプロジェクトを進めている。対象となる邸宅は、大正期の文化住宅の流れを汲む建築物で、建築史的に貴重な文化資源である。この坪井邸は、新築からその後の増築に関しての設計図・見積書・領収書・請求書などの 100 点余りの貴重な書類が残されている点に特色がある。これらの建築資料は当時の住宅建設がどのような過程で建設されたかを知る上で極めて貴重な資料であると考えられる。そのため、プロジェクトチームではこれらの資料をデジタルアーカイブ化することにした。しかし、これらの建築資料には、専門的な用語や、建設に関わったさまざまな組織、人名などの固有名詞が多く記載されているため、これらの資料をよりよく理解するためには、書類で使用されているタームの概念を体系化して、その関係性を明らかにする必要がある。そこで、プロジェクトチームでは、書類内で使用されているタームを切り出して分析し、建築資料のオントロジーを構築することにした。

オントロジーとは、さまざまな知識情報を階層構造に基づいて体系化することである。現在では、法律、産業、医療などのさまざまな分野で、各概念や用語を体系的に構造化する手法として使用されている¹。しかし、歴史的な資料を扱ったデジタルアーカイブやデータベースにおいて適用された例は少ない。日本においては、史料の目録データベースにオントロジーの概念を導入する例が先例とし

¹ 東京大学大学院学際情報学府² 東京大学大学院情報学環³ 東京大学大学院情報学環⁴ 溝口理一郎：「開発されたオントロジー」、人工知能学会誌, vol.19, No2, 2004

てみられるが、現状ではそれ以外の実践例が少なく検証が不十分であると考えられる⁵。そこで、プロジェクトチームでは坪井邸の建築資料を対象にしたオントロジーを構築し、それらがデジタルアーカイブにどのような有用性があるかについて検証することにした。

2. オントロジー構築作業

坪井邸の建築資料に関するオントロジー構築にあたり、プロジェクトチームでは以下の3つの手順で作業を行った。

- 1：資料に使われているタームを切り出す。
- 2：各タームを分析し、その背景にある意味や概念を分析する。
- 3：2の考察に基づいて建築資料に内在する概念体系を階層化し、オントロジーを構築する。

プロジェクトチームでは、上記の手順の中でまず始めにフェーズ1の作業に着手した。フェーズ1の作業はまず、建築資料で使用されているタームを切り出し、どのようなタームが使用されているかを把握する作業である。対象となる建築資料の記述内容は、書類のタイトル・宛名・価格・但し書きなど短い単語・文章で構成されている。そのため、これらの短い記述から名詞と動詞を中心にタームを切り出すことにした。

タームの切り出しを行った後は、フェーズ2と3に移り資料内部の知識情報の概念を体系化して、オントロジーを構築する作業を行った。その結果、後掲表1が示すようなオントロジーツリーが構築された。表1が示すように、オントロジーツリーの上位階層では「場所情報」「構造体情報」「個人・組織情報」「行為情報」「単位・値情報」「資料体情報」の6つの主なカテゴリで構成されている。まず、「場所情報」は建築資料に記述されている関係者の住所など、場所に関するタームがその範疇に入る。次に、「構造体情報」は坪井邸内の「応接間」「書斎」「屋根」などの家の構造内のある部分を示すタームがその範疇として入る。さらに、「個人・組織情報」は、建築に携わった建設会社、大工、職人など、個人・組織に関するタームがその範疇に入る。また、「行為情報」は各種工事や、事務作業などの行為に関する情報がその範疇に入る。そして、「単位・値情報」は、各種部材の長さ、価格、個数などに関する単位とその値の情報がその範疇に入る。最後に、「資料体情報」は、領収書、請求書などの書類や、各種建築部材、建築材料などの資料についての情報がその範疇に入る。オントロジーツリーは主にこれらの6つのカテゴリで構成され、それらの下位にそれぞれの書類で使用されていたタームがインスタンスとして属するようにした。

3. デジタルアーカイブにおけるオントロジーツリーの活用

前節において作成されたオントロジーツリーをデジタルアーカイブにおいて主に3つの機能に活用した。まず、第一の機能は検索における概念検索の導入に用いた。概念検索とは、オントロジーツ

⁵ 相田 満：「目録データベースの高次化によるデータマイニングを可能とするために—複数種のオントロジー辞書の利用・接合により検索効率の向上を試みる—」、情報処理学会 人文科学とコンピュータシンポジウム 論文集 2004 pp.151-158

リーにおける上位概念を予め検索画面にプルダウンメニューで示し、利用者はその上位概念をメニューから選択し、その下位に属するタームが使用された資料を検索することができる方法である。この検索方法では実際に使用されているタームを知らなくとも、上位の概念レベルで、該当する資料が検索できる。例えば、さまざまな木材の部材名が記述されている証書を検索したい場合、資料には「木材」といった概念レベルでの分類名が記述されているわけではない。そのため、通常であれば、「檜」「杉」「松」などさまざまな種類のタームをそれぞれ検索で指定する必要がある。しかし、概念検索を用いれば、用意されたプルダウンメニューから「木材」というクラス名を指定することで、その下位に属するさまざまな木材の部材名を記述した資料を検索できる。このように、具体的な単語を知らなくとも各分類を選択することで、関係する資料を検索することができる。

次に、第二の機能としては、オントロジーマップをアーカイブ上に設け、概念ツリーから直接各階層のクラスやその下位に属するインスタンスを選択できるようにした(図1)。何らかのタームを選択すると、各ターム別の解説画面が表示され、各タームの意味、そのタームが掲載されている資料、各タームと関連するサイトへのリンクなどの情報が掲載されるようにした。また、各タームの上下階層のクラスも表示されるため、この画面からオントロジーツリーの階層を上下、または左右に移動することができる。利用者は、この画面からターム同士の概念的なつながりを追って複数の資料画面へアクセスすることが可能となる。また、ターム解説の画面から資料が掲載されている画面にアクセスすると、各画面にはデジタル画像化された資料の他に資料内の文章を翻刻したテキストが表示されている。翻刻されたテキスト内の主要タームにはそれぞれリンクがつけられており、該当する各タームの解説画面へリンクがつけられている。

さらに、第三の機能としては、オントロジーツリーの中でカテゴリー化した6つの主要クラスの中で、「場所情報」「構造体情報」「個人・組織情報」「単位・値情報」の4つの観点から各資料の相関性を図で表し、各相関図よりそれぞれの資料にアクセスできるようにした(図2)(図3)(図4)。利用者はこの相関図から各資料の関係性を可視的に捉えながら各資料にアクセスができるようになっている。

これらの3つの機能により、利用者は、実際に資料に記述されているタームを知らなくとも、さまざまなレベルの概念を指定することで検索ができるようになり、掲載されているタームの意味的な関係性からさまざまな資料にアクセスすることができるようになった。さらに、オントロジーツリーの基本カテゴリーを可視的に捉えた相関図を設けることで、さまざまな資料の関係性を視覚的に確認しながら資料にアクセスすることができるようになった。

4. まとめと今後の課題

以上のような方法で、オントロジーを使用したアーカイブを構築した。オントロジーを使用し、建築資料内のタームの概念を階層化することで、概念レベルでの検索が可能となった。また、資料を翻刻して、そのテキスト部分を各ターム解説やオントロジーツリーと結び付けることで、各資料の各タームがどのような意味を持ち、資料全体の中でどのような位置づけにあるものかを確認しながら資料にアクセスすることも可能となった。さらに、オントロジーツリーの中の主要な概念を軸にし、各資料相互の相関関係を可視化することによって、各資料の関係性を基にそれぞれの資料にアクセスする

ことができるようになった。このように、専門的かつ、さまざまな関係性を持つ資料のアーカイブ化に際して、オントロジー構築とそれを利用した検索機能を付加することは利用者の情報アクセスにとって極めて有益であると考えられる。

ただし、建築資料のアーカイブの観点からみると、今回のオントロジー構築は建築資料全体の特性に目を向けながらもあくまで坪井邸の建築資料の特性に沿った体系である。そのため、これらのオントロジーツリーが他の建築資料にも適用できるかは今後検討する課題である。また、歴史的な資料などを扱うアーカイブ全体からみると、このプロジェクトにおいては、領収書や、請求者など、宛名・宛先・但し書き・価格などの単純な項目のみ記述されている資料であったが、より長い文章が掲載されている資料などにもこのようなオントロジー構築が可能であるかも今後検討すべき課題であると考えられる。これらの課題に関しては、今後さまざまな例証を積み重ねる必要がある。

オントロジーツリー

資料解説画面

ターム解説画面

図1: オントロジーツリーとターム解説、各資料画面とのリンク

図 2:「場所情報」に関する資料関係図

建築資料に記述されている関係者の住所と各資料との

図 3:「数値・単位情報(時間)」に関する資料関係図

建築資料に記述されている関係者の各資料と時間との関係性を可視化した。

図 4:「個人・組織情報」に関する資料関係図

建築に携わった建設会社、大工、職人など、個人・組織と各資料との関係性を可視化した。

表 1: 坪井邸建築資料に使用されるオントロジーツリー