デジタルネットワークにおける歴史的人名・組織情報の現状と その利活用モデル

Current state and the usage model of the information
about historical figure on digital network.

研谷紀夫
Togiya, Norio

*1 情報知識学会 所属 東京大学大学院情報学環

Japan Society of Information and Knowledge, Interfaculty Initiative in Information Studies
the University of Tokyo

E-mail: togiya.norio@iii.u-tokyo.ac.jp,

インターネットを中心とするデジタルネットワーク空間においては、電子化された歴史資料など多様な文化資源情報が格納されている。これらの情報内には様々な歴史的人名・組織情報が含まれているが、各人物に関する参照情報を提供することによって資料情報のより深い理解を促進させることが可能となる。本研究では、現在のインターネット上の人名・組織情報の現状を調査した上で、デジタルネットワーク上で共用可能な主に明治以降の近代期を対象とした歴史的人名・組織典拠情報の可能性について検討する。

There is currently a variety of computerized historical and cultural resource information stored on digital networks, namely the Internet. This information contains a variety of information on historical figures and through the provision of such reference material it is possible to promote a further understanding of historical and cultural archives and documentation. This research considers the potential uses of authentic data concerning historical figures in the modern age that can be shared over digital networks following examination of the current state of such information available over the Internet.

キーワード: 人名典拠, 文化資源, アーカイブ, Authority Control, Cultural Resource, Archive

1. 本研究の背景と概要と背景

インターネット上には様々な歴史的情報やデジタル化された文化資源が格納されている。これらの情報の中で重要な情報項目の一つが、歴史的な人名・組織情報である。このような人名・組織情報は歴史資料などに記載される人名・組織名その他、文化資源の作成・寄与者、歴史的な出来事と関わりのある人名などがそれらに対応する。

しかし、これら歴史的な人名・組織名につ

いては、同名異人、異名同人などが存在し、検索精度の向上のためには各存在に基づいた名称の統制が必要である。人名・組織名の情報コントロールについては、図書館情報学の分野での実践が先行し、日本では「国立国会図書館人名典拠録」などが知られている。

これらに対し、日本では拙論[1]で指摘したように、自然科学系博物館、美術館、文書館などで使用される人名典拠情報の整備実績

は少数である。このような状況の中で、インターネットを中心とするデジタルネットワークの普及によって図書、博物館、文書館の枠を超えた情報の共有が進み、これらを包括的に扱うことが可能な人名・組織典拠情報の整備が必要となっている。

本研究ではこれらを踏まえ、明治時代以降の近代期を対象に、インターネット上の文化情報資源で用いることが可能な人名典拠データのあり方を、その運営主体や情報内容、情報の再利用などを含めた全体像を提示する。著者は既に図書館、博物館、文書館などで共通に用いることのできる、具体的なエレメントセットなどについて提案しているが[1]、本論文では検索エンジンでの利用なども踏まえた、視野を広げた全体的な事項について考察する。

本来ならばこのモデルに基づいた実践を行った上で、研究成果を提示することが望ましい。しかし、考察の範囲が社会・文化的な文脈を踏まえた広範囲な事柄を含むため、本稿では現状調査とそれらに基づいたモデルを示した上で、今後の取り組みにおける課題と展望を示す。

2. 人名・組織典拠情報のあり方

デジタルネットワーク上で活用する典拠情報に関しては、これまで図書館を中心に構築されてきた典拠情報と違い、様々な文化資源を対象とする、より幅広い名称の統制が必要となる。これらに関しては、芸術・文化、学術・研究者、政治、産業など様々な分野の人名及び組織をあげることができる。また、組織においても同様に、図書を中心とした人名典拠情報においては出版社など書籍と関わりの深い組織名の統制に限られていたが、政府・政治組織、役所・官庁、民間企業、文化団体・流派など、より広い文化情報資源を対象にする必要がある。

これらの人名・及び組織の名称に関して特に統制が必要となる項目は、表-1 のような諸

課題があげられる。まず始めに人名のみを対象とすると、最初の課題として同名異人をあげることができる。そして次には、同一人物が時間的経緯に基づいて名前が変化していく場合や、同時に複数の種類の名前を持つ異名同人をあげることができる。またさらに、常用漢字、旧漢字、俗字の使用などから生じる異表記で同一人物を指し示す課題などがある。

一方で組織については、同名異組織と、異名同組織などの課題がある。後者に関しては、歴史的経緯において名称が変わる場合と、正式名と通称・ブランド名といった同時期に複数の名称・通称が生じる場合が存在する。また組織においても人名と同様に、常用漢字、旧漢字、俗字の使用を原因として、異なる表記で同じ組織を指し示す課題をあげることができる。

これらの中で、特に組織に関しては本研究で対象とする近代以降の経済の発達による法人組織の急増に伴い、同じ企業名が多く存在する。そのため人名とともにその適切な同定可能なデータベースが産業史などを行う上で重要となる。しかし、現状においてはこれらを歴史的観点から包括的に集約した無償の組織名データベースが存在しない。そのため、これらの情報を掲載した典拠データベースが必要である。

3. 現状の人名・組織情報

3-1. 検索エンジンの対応

本項では前項におけるような人名・組織名の統制について検索エンジンや各データベースの辞書機能の現状について分析する。最初に検索エンジンやインターネット空間全体を対象とする文書検索エンジンなどにおける現状について考察する。

これらの分野においては、人名に関係する固有表現の抽出と、各人名が掲載された文書の文書間の類似度計算によって、各同名異人の識別を行う手法が最も有効な方法

であると考えられる[2]。しかしこの方法はコンテンツの量と質に依存している面が強く、ネット上で叙述が無い歴史的人名・組織名などに関してはその識別や他との関係を導くことは難しい。

さらに、同名異人ではなく異名同人に関する識別に関しては、特に戦前期の人物は生涯の中で多くの名前を持っていることが多く、分析の対象となる文章に十分な名称情報が記述されていない場合が多い。

さらに構造化されていない文章からの分析では、十分な情報が存在していてもそれら複数の名称相互の関係を把握することは困難である。そのため、歴史的な異名同人及び同名異人の識別精度をあげるためには、より根本的な対応が必要であろう。

3-2. 公共データベース

これらに対して、より精度の高い人名情報を統制してきたのは、図書館、文書館などの公共機関で作成された典拠データベースである。日本においては、国立国会図書館や国立情報学研究所などの人名典拠情報が知られている。しかし、前述したようにこれらの対象は著作や編集などを行った人物であり、広く歴史上の人物を扱っている訳ではない。そのためこれらを利活用しつつも、より他分野の人名典拠情報を構築する必要がある。

また、近代以降の歴史上の人物に関しては、国立公文書館や付属するアジア歴史資料センターなどで構築されているデジタルアーカイブの辞書機能をあげることができる。これは、人名や組織名とそれらと関連する用語を格納した辞書である。これらは必ずしも人名や組織名に特化した辞書ではないが、異名同人に関する識別には一定の有効性がある辞書機能である。

さらに、国文学研究史料館では、「国書総目録」「日本人名辞典」などを基礎とした日本の「古典人名データベース」がある[3]。これら

は古典を対象としているため本研究とはその時代対象がやや異なっているが、そのエレメントセットや内容記述の課題については示唆に富む事例である。これらの公共データベースにおいて蓄積された人名・組織名情報は質・量ともにデジタルネットワーク全体の人名・組織名統制にも極めて有用なデータである。

しかし、これらはそれぞれの分野や特定の時代の情報に特化していることや、同じ構造化手法でデータが統制されていない点、さらに全てが無償で一般に公開されているものではなく、購入が必要であることや、他のデータベースでの利用が難しい面が多い。そのため、これらの成果を基礎とした、よりデジタルネットワーク上で誰もが利用できる人名・組織典拠情報の構築を模索する必要がある。

3-3. 民間・個人データベース、書籍の概要

一方、公共機関とともに、非営利/営利の民間法人においても数多くの人名・組織名を掲載したデータベースが存在する。その代表的な例として、オンライン百科事典であるWikipediaをあげることができる。同サイトでは対象とする人物によって記述の差はあるが人名の他、生没年と略歴などが掲載され、同名異人が存在する場合は「曖昧さの回避」の情報によって、ナビゲーションがなされている。また同記事中に時間的な変遷によって名称が変化する場合は、その概要が解説されている。

しかし、これらのデータは共通のフォーマットによって構造化が行われておらず、内容によっては他のコンテンツより剽窃の疑いのある記事や内容記述によって論争が起きている記事などもある。そのため、参考とする基礎的な情報として活用することが望ましい。

また、個人のWEBページなどにおいても著作権フリーの人名・組織辞典などが多数存在する。これらに関しても情報の精度などを

確認した上で、参考情報として活用できる面においては管理者の許諾を得た上で活用し、これらの資源を公共化し持続可能な形で継承することも考えられる。

また、これ以外にも民間企業による有料・無料の人名・組織データベースなどが存在し、図書の分野においては「TRC MARC 人名典拠録」などが知られている。また組織名などに関しては民間のデータベースがより豊富なデータを用意している。しかし、これらの情報はデジタルネットワーク上で無償に利用するデータとして利用することは困難である。

これらから同名異人、異名同人などの問題が生じる人物などについては、権利問題に抵触しない限りにおいて、公共機関や非営利法人、個人のデータベースを参考とし、書籍などの信頼性の高い人名辞典で確認を行いながら、より広い枠組みの人名典拠データベースを構築することが望ましい。そしてその上で、無償でそのデータを提供し、デジタルネットワーク全体の人名・組織名の判別に寄与することが望ましいと考えられる。

4. 課題へのアプローチ

4-1. 人名典拠情報モデルの概要

前項において考察した現況を踏まえた上で、デジタルネットワーク上で共通に使用できる近代人の人名典拠情報について考察する。人名典拠情報に関しては最初に述べたように、同名異人、異名同人の統制を第一の目的として、これらの問題が生じる人名・組織名を扱うことを大きな目標とする。また日本語だけを対象とするのではなく、日本の人名・組織名の英語表記についても対応することが望ましい。特に、組織においては英語表記が日本名と若干異なる場合があるため、近代以降の産業資料などと対応させる意味では重要な検索ツールとすることが必要である。

4-2. 人名・組織名典拠情報モデルの内容

第二項であげた人名典拠情報の役割と、第三項であげた現在の状況から、デジタルネットワーク上で構築する人名典拠情報の概要を設計すると、具体的には、大きく分けて「A. 活用目的・活用モデル」「B. 記述対象・内容・方法」、「C. 他の人名・組織情報との関係」、「D. 社会的要素」の主に四点によって構成される。それぞれの概要は以下の通りである。

A. 活用目的・活用モデル

- ① インターネット上のデジタルミュージアム、アーカイブ、ライブラリーなどにおけるデータ検索時において、近代期の同名異人(組織)・異名同人(組織)などが生じる日本の人名・組織名の日本語、英語両面の統制に用いる。
- ② 各種の検索エンジンなどにおいて、歴史的固有名詞の辞書の作成や、ユーザ検索時において、近代期の同名異人(組織)・異名同人(組織)などが生じる人名・組織名の日本語・英語両面の統制に用いる。

B. 記述対象・内容・方法

- ③ 歴史上同名異人(組織)・異名同人(組織)、表記の問題が生じる日本語・英語両面の人名・組織名を対象とし、それらの情報統制情報を格納する。
- ④ 対象は近代期の人名・組織とするが近代以前の人名などにおいても対応できるフォーマットを目指す。
- ⑤ 記述内容はその存在が明らかになっているもの限定し、中立的かつ客観的な情報を掲載する。
- ⑥ 対象は、博物館、図書館、文書館などでも使用できるよう、芸術・文化、学術・研究者、政治、産業などより広いジャンルで活躍した人名や関連する組織を対象とする。
- ⑦ 名称及びその記述には英語での解説な

どを取り入れる。

- ⑧ 主要な掲載内容としては、ID、名称、別名(俗称・通称、略称、雅号)、生没年、略歴、引用資料や参考サイトなどへのリンク情報を掲載する。
- ⑨ その他の人名との社会的関係性を記述する。
- ⑩ IDは統一的なIDの管理と、他の人名典拠データのIDを掲載する。
- ⑪ WEBオントロジ言語など機械可読式言語で記述する。

C.他の人名・組織情報との関係

- ⑫ 権利面の問題をクリアした上で、現在存在するデータベースを活用し信頼性の高い書籍の情報で確認を行う。
- ⑬ 様々なデータベースでの包括的なデータの利用を可能とする。
- ⑭ 事後的な編集を可能とするインターフェイスを提供する。
- ⑮ 様々な人名典拠情報とマッピングを行える仕様とする。

D.社会的要素

- ⑯ パブリックドメインの情報として活用する。
- ⑰ 持続可能な組織で運営する。

これらの概要について詳説すると、人名・組織に関する典拠情報は、⑰で示されたような、持続可能な組織の中で、⑤の中立的かつ客観的な視点において近代期の③同名異人(組織)・異名同人(組織)に関する統制情報を掲載し、⑯の視点に基づき広く無償で公開することを第一の目的とすることが望ましい。これらの情報を内外の検索エンジン、様々な文化資源を扱う図書館、文書館、博物館・美術館のデータベースやデジタルアーカイブに用いることによって、デジタルネットワーク上の文化資源の検索の利便性を向上させることが目的である。

またそれらを実現するためには、⑩で示すように、統一的なIDの管理と、他の人名典拠データのIDが掲載されていることが必要である。さらに、⑪の機械可読式のオントロジ言語で記述することによって、コンピュータプログラムがそれらの名称情報を構造的に把握することや、英語などのサイトでも利用できるように、⑦英語での名にも対応することが必要である。また、⑨と⑧が指摘するように、他の人名組織名との関連を記述するとともに、引用資料や参考サイトなどへのリンクを設置することが必要である。

さらに、これの実現のためには、無からデータを構築するのではなく、⑫と⑬にあるように、権利問題に抵触しない限りにおいて、現在ある情報を可能な限り活用し、様々なデータベースでの包括的に典拠情報データ利用を可能とすることが必要である。また対象とする人名は近現代期であるが、④が指摘するように、それ以前の時代の人名・組織名にも対応可能な設計を行う。さらに今後の運用面においては、⑭と⑮にあるように、事後的な編集を可能とするインターフェイスの提供や、その他の様々な人名典拠情報とマッピングがとれることが重要である。

5. 今後の課題

前項で示した概要がデジタルネットワークで広く用いることが可能な人名・組織典拠情報の概要である。しかし、実際の記述には近代期の人名・組織においても、相田が指摘するように、エレメントセットのデザインや、時代とともにその扱いが変化する姓、諡号、諱などの扱い、同一人物でのよみの多さ、さらに俗字、合成文字、梵字の扱いなどの課題が存在すると考えられる[3]。

これらの人名データベースの課題はデジタルネットワーク上の人名・組織名典拠データベースにおいても生じる課題である。そのため、これらの構築には、既存の図書館、文書館、博物館における人名典拠情報の現況

を調査し、そこで得られた資産や知見をデジタルネットワーク上で再活用する視点が望ましい。その上で、これらの新しいネットワークでの活用の方法に合わせた、典拠情報の設計と構築が必要である。

参考文献

[1] 研谷紀夫, ネットワーク文化情報資源で活用する人物典拠情報の構築とその課題, 第55回日本図書館情報学会 研究大会発表要項, 2007, pp73-76

[2] 小野真吾 他, 重要語抽出を用いたWeb 文書上の同姓同名の曖昧さ解消, 電子情報通信学会 第19回データ工学ワークショップ予稿集, 2008,
<http://www.ieice.org/~de/DEWS/DEWS2008/proceedings/files/a7/a7-3.pdf>
(2008.04 Accessed)

[3] 相田満, 古典人名データベース作成上の問題点, 情報知識学会研究報告会講演論文集, 1993, pp27-32

表-1: 同名異人(組織)、異名同人(組織)の代表的な例

A: 人名	
A-1-a: 同名異人	
吉田茂	外交官出身の政治家。戦後第45代の内閣総理大臣を務める。
吉田茂	戦前期の内務官僚出身の政治家。貴族院議員、軍需大臣、厚生大臣を歴任。
A-1-b: 同名異人だが読みが異なる場合	
早川徳次(はやかわとくじ)	電子メーカー「シャープ」の創業者
早川徳次(はやかわのりつぐ)	日本初の地下鉄である東京地下鉄道の創業者
A-2-a: 異名同人(歴史的経緯)	
木戸孝允	1865年からの名前
桂小五郎	1840年頃から実質的に使用された名前(他にも多数の名前・通称が存在する)
木戸準一郎	1866年頃からの名前
A-2-b: 異名同人(同時期に2つの名前が存在する場合)	
桂小五郎	1840年頃から幕末までの実質的に使用された名前
新堀松輔	幕末頃の変名
松菊	正式名とともに用いた雅号(雅号は他にも多数存在する)
その他役職名、あだ名、通称などがある。	
A-3: 同名同人だが、表記が異なる場合(主に旧漢字・俗字と常用漢字など)	
山縣有朋	旧漢字
山県有朋	新漢字
B: 組織	
B-1: 同名異組織	
東京電力	1951年に創設された電力会社
東京電力	1925年早川電力、群馬電力が合併してできた会社。現在の東京電力とは別会社
鈴木商店	1874年創設の戦前の財閥。世界恐慌時に倒産した。
鈴木商店	1912年に鈴木製作社から改名した味の素の前進の会社。財閥の鈴木商店とは別会社。
B-2-a: 異名同組織(歴史的経緯)	
箕面有馬電気鉄道	1906年創設の鉄道会社
阪神急行電鉄	1918年変更の社名
京阪神急行電鉄	1943年に京阪を合併して誕生(戦後分離)
阪急電鉄	1973年に社名変更
B-2-b: 異名同組織(同時期に2つの名が存在・混同されている場合)	
金鳥	会社名とも誤解されやすいブランド名
大日本除虫菊	正式会社名
大阪電気軌道	大正・昭和期に存在した近畿日本鉄道の前身にあたる鉄道会社の正式名
大軌	大阪電気軌道の略称・通称
B-3: 同名同組織だが、表記が異なる場合(主に旧漢字・俗字と常用漢字など)	
文藝春秋社	旧漢字を用いた正式社名
文艺春秋社	常用漢字を用いた会社名