

人名典拠情報の API による共有化と 画像資料における人名情報参照システム

研谷紀夫 川島隆徳

東京大学大学院 情報学環 東京工業大学大学院 社会理工学研究科

本研究では博物館、図書館、文書館で利用できる人名典拠情報を作成し、それらを API サービスで提供を行う。このサービスを用いることによってそれぞれの Digital Cultural Heritage は歴史的な人名や芸術家に関する統合的な人名を用いることが可能となる。本稿では実際に、これらの API を戦前の写真をデジタル化して格納した Digital Cultural Heritage で活用した例について解説を行う。

A System for Referencing Personal Names through Iconography and Sharing an Authoritative Information Source for Personal Names by API

Norio Togiya

Interfaculty Initiative in Information Studies
University of Tokyo

Takanori Kawashima

Graduate School of Decision Science and Technology
Tokyo Institute of Technology

For this study, a usable authoritative information source for person names from museums, libraries and archives was created and provided through API services. By using this service, it is possible for respective Digital Cultural Heritages to have a comprehensive name list of historic personal names and artists. These API are used for Digital Cultural Heritages of digitized old photographs, and old photographs can be used for analysis. In this paper we will present a summary of these.

1. まえがき

写真などを含めた画像資料が歴史や文化研究などの対象として用いられるようになって久しい。しかし、文字による情報が記載されていない画像情報については、被写体に関する詳細な情報と組みあわせた調査と分析が必要である。そのため、これらの情報を効率的に取り入れて、比較・検証するシステムが必要である。

これらの情報の中で、最も使用頻度の高い情報の1つが「人名情報」である。特に様々な機会に際して撮影された記録写真には多くの人物が撮影されており、これらの人名を特定することによって、撮影された時期や被写体となった出来事や場所などを分析する重要な手がかりとなる。そのため、これらの人名とのシームレスな連携を行うことが望ましい。

本研究ではこれらの観点から、戦前の華族に関する人名典拠情報を格納する。その上で、これらの情報を、API を通じて分析ツールで利用し、写真資料の分析に用いた事例について解説する。

2. 典拠情報を利用した画像分析の全体像

本研究の特色は、日本においては研究の対象としてあまり注目されてこなかった人名典拠情報に着目し、文化資源をデジタル化した Digital Cultural Heritage などのサービスで実践的に活用することを主眼とした点である[1]。特に API による典拠情報の提供は実践例がほぼ存在せず、これらを外部のデータベースに提供していくことを目標としている。

具体的な典拠情報としては、著者が既に蓄積した戦前の政・財・軍・官で中心的役割を果たした華族と、写真師や画家などを中心とした人名典拠情報を対象としている。

これらの典拠情報は、図書館、文書館、博物館で共通に用いることが可能で、ネットワーク文化資源でも活用できる典拠情報の設計と格納を目標に蓄積がなされた[2][3][4]。

蓄積された典拠情報をより有効に活用する手段として、典拠情報のデータを、API を通じて供給する“ウェブサービス”として流通させることが本研究の柱である。

文化資源情報の API の提供については日本の国立国会図書館が、目録データベースの検索 API を提供しているが、まだその実践は一部であり、情報の更新・編集までは行われていない。そのため本研究では、典拠情報を対象として他の Digital Cultural Heritage で活用する事例に関する実験を行った。

また、これまで蓄積した人名情報は相互に血縁、地縁、同門、同窓、同期などによって様々な関係性で結ばれている。そのため、それぞれの人間関係のつながりから、関連する資料の系統・分類化にも活用できると考えられる。本研究では、このように様々な可能性を持つ人名典拠情報をウェブサービスの一環として捉えるとともに、それらを資料の系統・分類化に用いるなど、新しい活用方法を模索する研究である。本論文ではその中で、蓄積された典拠情報を歴史的な写真が蓄積された Digital Cultural Heritage において活用する実験について解説を行う。

3. 典拠情報を利用した画像分析の全体像

本研究においては、図1で示すように、人名に関する典拠情報データベースを構築し、それらの情報を、API を通じて画像資料の被写体分析システムで分析に用いる構成となっている。

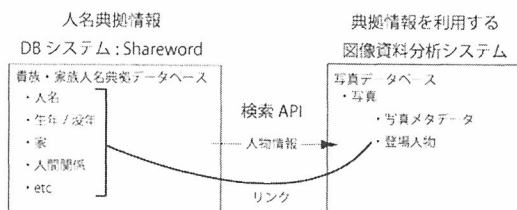


図1. システム全体像

人名典拠情報については、戦前の皇族・華族と写真師を中心として、約 3000 名の人名情報を格納し、DB システムである Shareword に格納している[8]。これらの人名典拠情報は、図書館、博物館、文書館で用いられる標準的な人名典拠形式に主要な部分で対応し、情報の共有などが可能なエレメントセットで作成されている[9]。本研究では、このデータベースを外部から検索するための API を追加した。さらに、その API を利用する外部システムとして写真アノデータを試作した。

4. 人名典拠データベース

人名典拠情報は、典拠情報データベースシステムである Shareword に搭載されている。全 282 家、3415 人の皇族・華族の典拠情報と、人物間の血縁関係 3524 関係が格納されている。典拠情報は表 2 に示される全体のデータの中から管理情報などを除いた表 1 の内容となっている。

表 1. 典拠情報フィールド

代表的な名称	人物の代表的な名前、その読み
その他の名称	その他の名前と読み
略歴	人物の略歴
生年	生年月日、あれば元号
没年	没年、元号等
関係 URL	wikipedia のエントリーなど、関連する URL
所属グループ	人物が所属したことのある集団
役割	人物がついたことのある役職
出典	データの出典

人物及び家族にはそれぞれ UUID (Universally Unique Identifier, ISO/IEC 11578) が割り振られており、この ID を用いることで同姓同名の人物であっても一意に識別することが可能である。

5. 人名典拠 API の仕様

本研究では上記の典拠情報 DB から人物情報を検索、取得するための API を設計し実装した。この API を利用することで、外部のシステムでも統制された人名情報を利用することが可能になる。

API の形式としては、近年 Web 上の分散システムで利用されている REST [10] のアーキテクチャを採用した。REST では、URI を用いて Web 上のリソースを指定し、HTTP の GET を用いてリソースを取得する。本システムの場合、リソースは人物の典拠情報にあたり、XML を用いて表現される。

検索に用いることが可能なパラメータは {人物 ID, 人名, 家 ID, 家名} の 4 種類である。人名に関しては、代表的な名称やその他の名称、さらに、それぞれの読みの部分も検索される。検索はこれら 4 つのパラメータの AND で行われるが、人物 ID を指定した場合には一意に人物が決定されるため、他のパラメータは無視される。また、家 ID を指定した場合、家名のパラメータは無視される。

検索によって取得されるデータは、{人名, 人物 ID, 生年, 没年, 家名, 家 ID, 略歴, 関連人物とその関係, 出典} の 8 項目である。

具体的な検索は、検索用 URI を用いて行われる。例えば、名前に山田を含む人物を検索する場合、「http://ホスト名/noble?name=山田」と

いう URI を HTTP で GET することにより、検索結果の一覧が XML として返ってくる。XML の一例を図 6 に示す。このような REST の API は、ブラウザだけで検索することも可能であるし、HTTP と XML を扱えるシステムであれば利用可能で、AJAX などによるマッシュアップも容易である。

なお、現段階はサーバを公開するフェーズではないため、Web 上で検索サービスの提供は行っていない。

6. 典拠 API の利用

次に典拠情報を API を通じて提供するために、以下の二通りの利用方法が考えられる。

(1) 典拠情報の ID を用いて、人物を一意に識別する：この利用方法は、本来典拠情報が果たすべき役割であり、本 API の本質的な機能に当たる。主に「名前」によって定義される人物であるが、本 API を利用して名前で検索し、その人物の ID を取得することで、揺らぎのある表記や、同姓同名の区別といった問題を回避することができる。さらに、この人物 ID を複数のシステムが共有することで、システム間の横断検索や連携が可能になる。例として、ID を共有した異なる機関の DB を横断検索する手順を図 2 に示す。

(2) 典拠情報に記されたデータを閲覧する：この利用方法では、単純に API を通じて人物を検索し、その情報を取得することを目的とする。この目的は副次的であるため、現時点では API は名前や家以外のフィールドを用いた検索を提供していない。

本研究では、この 2 つの利用方法を実証するべく、2 つのシステムを開発している。(1) の利用方法に当たるのは、写真のアノテータであり、(2) の利用方法に当たるのは典拠情報ブラウザである。

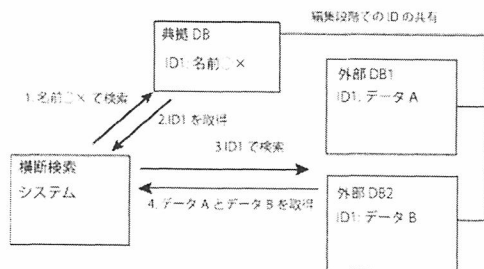


図 2. 横断検索システムの一例

7. 写真アノテータ

写真アノテータは、華族等の集合写真に誰が写されているのかアノテーションを付与し、それに基づいた解析を行う。実験用アプリケーションとして、Java によるスタンド・アローンのアプリケーションとして実装した。機能は大きく検索、アノテーション、解析の 3 つに分けられる。

(1) 検索：検索機能では、アノテーションをつける写真を検索することができる。それぞれの写真には、タイトル、作者、作成年月日、出版、関連組織、場所、時間、権利のメタデータを付与することができ、基本的にはそれらを用いて検索を行う。また、人物アノテーションが付与されている写真であれば、登場人物を指定して検索することも可能になる。

(2) アノテーション：図 3 はアノテーションの画面を示している。アノテーションは、人物の顔を○で囲み、その人物の名前を付与することで行う。名前は典拠 API を利用して典拠情報の人物と関連付け、アイデンティファイする。それにより、写真上で典拠情報がポップアップで表示され、後述する血縁関係の検索等が可能になる。現時点では、典拠情報に存在しない人物は、ローカルデータベースに保存されるだけで、特定はなされない。これは、統制が必要な典拠情報に容易にデータを追加させないための仕様であるが、今後新規データ追加用の API を導入することも検討している。

(3) 解析：発展的な機能として解析機能を付与した。解析機能では、付与されたアノテーションから人物間ネットワークのグラフを形成させる(図 4)。このグラフは、各人物をノード(点)、同じ写真に登場する人物の間にエッジ(線)を張ったもので、写真での同時出現から見た人間関係を可視化することが可能である。さらに、典拠 DB に格納された血縁関係のデータを API を通じて取得し、特定の二人の人物の間にどのような血縁関係があるのかを検索することができる。これら二つの機能を組み合わせることで、いつも同じ写真に写っている二人の人物が、どのような血縁関係にあるのかを検索することが可能になる。



図 3. アノテーション画面

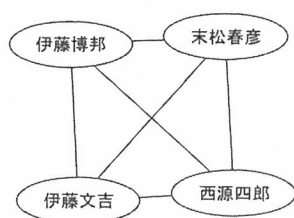


図 4. 人物出現ネットワーク図

8. 典拠情報ブラウザ

典拠情報ブラウザは、API による検索とその検索結果をユーザに扱いやすい形で提示するシステムで、Web アプリケーションの形態を取る。現時点では、人名、家名による検索と検索結果の一覧表示、そして関連人物や家族の一覧を表示するリンクを提供するだけの、単純なシステムである。これらのシステム群の概観は図 5 に示された通りである。

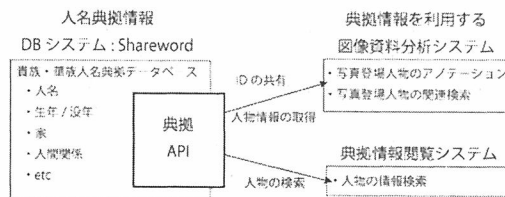


図 5. システムの関連図

9. 活用と今後の課題

本システムは、研究対象となる家に残る写真資料を調査し、それらの被写体となっている人物が誰であるか、またどこでいつ頃撮影された

ものであるかを調査していく際の研究支援ツールとなるものである。本研究では、旧華族であった複数の家に残る写真群の調査に利用している。写真の裏やアルバムのキャプションに書かれている人名情報を元に写真の人物の部分にアノテーションを付与し、人名情報を記録した。これらの人名情報は人名典拠データベースから API を通じて関連づけているため、当該人物の社会的な地位や姻戚関係を参照することができる。また、集合写真などについては、複数の人物に同じような作業を行うことで当主を中心とした姻戚関係の中で、どの部分までの親戚が集合写真に写されているかなどの基礎的な情報を取得する際の助けとなる。人名典拠情報は元データベースから API で提供されているため他のアプリケーションやデータベースでも用いることが可能である。

これらのツールを用いた写真などの画像資料の調査を行うためには、対象となる写真が、対象人物や家の成員の行動のどの程度を網羅しているか、あるいは資料全体の残存状況やその中のどの部分を占める資料であるかを客観的に把握した上で調査を進める必要がある。また、調査内容の精度を上げるためには他の資料との比較・検証が必要である。そのため、API の共有を通して、他のデータベースとの連携などを今後進める必要がある。

謝辞

本プロジェクトは科学研究費補助金「ウェブサービスによる典拠情報の提供とその活用に関する研究」などの協力を得て行われました。

参考文献

- [1] 宮澤彰, 典拠データに関する新しい取り組み, 図書館目録とメタデータ, 日本図書館情報学会編, pp23-36, 2004
- [2] 研谷紀夫, デジタルネットワーク上の文化情報資源で活用する人名典拠情報の JAPAN/MARC 形式への適応の試みとその可能性, 2009 年 3 月 TP&D フォーラム 2008 実行委員会, TP&D フォーラムシリーズ 整理技術・情報管理等研究論集 第 18 号, pp17-32, 2009
- [3] 研谷紀夫 馬場章, 文化資源統合アーカイブにおける時空間情報の活用とその課題, 情報処理学会シンポジウムシリーズ vol.2007, 人文科学とコンピュータシンポジウム論文集 2007, 情報処理学会, pp147-154, 2007

[4] 研谷紀夫,デジタルネットワーク上の文化情報資源で活用する人名典拠情報の JAPAN/MARC 形式への適応の試みとその可能性,2009 年 3 月 TP&D フォーラム 2008 実行委員会,TP&D フォーラシリーズ 整理技術・情報管理等研究論集 第 18 号,pp17-32, 2009

[5] 研谷紀夫 内藤求,トピックマップを用いた人名典拠情報の構築,情報知識学会,情報知識学会誌 Vol.19 No2,pp57-62,2009

[6] Norio TOGIYA,Motomu NAITOU,Construction of Authority Information for Personal Names Focused on the Former Japanese Nobility Using a Topic Map,Proceeding of TMRA2009,2009

[7] 研谷紀夫 川島隆徳,ネットワーク文化情報資源で活用する人名典拠情報とその共同構築のためのオントロジエディタ,情報処理学会シンポジウムシリーズ vol.2008,人文科学とコンピュータシンポジウム論文集 2008, 情報処理学会,pp113-118,2008

[8] 川島隆徳 研谷紀夫,著者名典拠情報を拡充するための共同編集プラットフォーム,情報知識学会誌,情報知識学会,20 巻 1 号 pp24-29,2010

[9] 研谷紀夫,社会的ネットワークを重視した人名典拠情報の構築 -戦前期の写真師を対象とする人名典拠を中心として-,アートドキュメンテーション研究,アート・ドキュメンテーション学会,17,pp31 -52,2010

ション研究,アート・ドキュメンテーション学会,17,pp31 -52,2010

[10] Roy Thomas Fielding,Architectural Styles and the Design of Network-based Software Architectures, <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>(参照 2010 -11-01)

[11] 三浦敬子,松井幸子. 欧米における著者名典拠ファイルの共同作成の動向. 日本図書館情報学会誌,vol.47, no.1, p.29-41,2001.

[12] Bennett, Rick. et al,バーチャル国際典拠ファイル(VIAF : Virtual International Authority File): ドイツ図書館と米国議会図書館の典拠ファイルを利用して, World Library and Information Congress: 72nd IFLA General Conference and Council. Seoul, 2006-08-20/24. http://www.ifla.org/IV/ifla72/papers/123-Bennett_trans-ja.pdf (参照 2008-12-01).

[13] 国立情報学研究所,日本語, 中国語, 韓国語の名前典拠ワークショップ記録, <http://www.nii.ac.jp/publications/CJK-WS3/> (参照 2008-12-01)

[14] 国立国会図書館,国立国会図書館著者名典拠の最近の動向について, http://www.ndl.go.jp/jp/library/data/010330_2.html (参照 2008-12-01)

```
<persons>
  <name>田中 信豊</name>
  <personID>a20a58ea-81ab-422c-a8e2-9f56e41e1075</personID>
  <houseName>田中</houseName>
  <houseID>4748415b-7110-444d-beb5-ea01cbd478e9</houseID>
  <birth>大正 1(1912/8/4)</birth>
  <history>町田孝豊二男, 初, 田中定孝養子</history>
  <relations>
    <personID>30a6ef76-073a-4c4b-89ae-6ee42dc27a03</personID>
    <personName>田中定孝</personName>
    <relation>子 (次期当主) </relation>
  </relations>
  <source>書名,頁数 </source>
</persons>
```

図 6. 検索結果の一例 (名前は仮称サンプル)

表 2: 管理情報などを含めた人名典拠の情報項目

項目	
名前 (複数可能)	(漢字・よみ・英字) (複数可能)
	世系 (複数可能)
	名前の種別 (別名雅号など) (複数可能)
国籍 (複数可能)	
性別 (複数可能)	
職業 (複数可能)	
人物 ID	
関係 URL/URI	人物 URI
	関係 URL (複数可能)
生没年	生年月日 (西暦・新暦のみ) (複数可能)
	没年月日 (西暦・新暦のみ) (複数可能)
略歴	略歴 日本語
	略歴 英語
誕生した地 (複数可能)	
存在した地 (複数可能)	
管理データ	入力日付 (複数可能)
	最終更新日付,
	種別
	言語コード (複数可能)
	文字コード
	典拠の確立状況
	入力者 (複数可能)
関係 (複数可能)	
	師匠・弟子・知人・父・母・兄・姉・弟・妹・夫・妻・子