

複数の異種浮世絵データベース間における同一作品の 同定手法の提案

久山 岳夫¹

Biligsaikhan Batjargal²

木村 文則²

前田 亮³

¹立命館大学情報理工学研究科 ²立命館大学衣笠総合研究機構 ³立命館大学情報理工学部

本論文は、複数のデジタル化された美術館・博物館の情報に対して統合的なアクセスを可能にするプロジェクトの一環について述べたものである。本研究の目標は、異なる形式や言語で記述されている複数の人文系のデータベースに対して同一レコードの同定を可能にする手法の提案を行うことである。本論文では、人文系のデータベースの一つとして、特に浮世絵の画像データベースを対象とした。本論文では、同一作品の同定を行うために、浮世絵のメタデータを提供している複数の異種データベースのレコードに対して、主にレコードのタイトルを用いて他のレコードとの類似度を算出する。これにより、複数のデータベース間に散在する、特定の浮世絵の類似作品を見つけること、また、どの組織がその浮世絵、あるいはその複製を持っているかを知ることが可能になる。提案手法を実装したシステムの同定精度の評価では、比較する二つのデータベースのレコードの記述が同言語の場合と異言語の場合に分け、比較を行った。同言語のレコード同士の比較では高い精度となったが、異言語のレコードの比較では固有名詞の辞書を用いることで同定精度を向上させることができた。

Identifying the same prints across multiple and diverse Ukiyo-e databases

Takeo Kuyama¹ Biligsaikhan Batjargal² Fuminori Kimura² Akira Maeda³

¹ Graduate School of Information Science and Engineering, Ritsumeikan University

² Kinugasa Research Organization, Ritsumeikan University

³ College of Information Science and Engineering, Ritsumeikan University

This paper provides the research achievements of our ongoing project in providing integrated access to multiple databases of digitized materials in libraries, archives, and museums. The main goal in this research is to propose a method, which allows identifying the same prints across multiple humanities databases in various different formats and written in different languages. In this paper, we utilize Ukiyo-e databases as an example of humanities databases. In order to identify the same records across multiple and heterogeneous databases that provide metadata of Ukiyo-e prints, we calculate the similarity degree of the given records using the data mainly in the title. Succeeding the proposed method, in other words, finding the same prints across multiple Ukiyo-e databases will help users to know which organization has a certain print and its variations. In the experimental evaluation, we analyzed the precision of the proposed method for two different conditions 1) identifying the same records in the same language and 2) in different languages. The proposed method achieved a high precision in identifying the same records in the same language. Besides, the precision of identifying the same records in different languages has improved by using a proper nouns dictionary.

1. はじめに

近年、美術館や博物館の所蔵資料のデジタル化が進んでおり、浮世絵も国内外の多くの機関においてデジタルアーカイブが公開されている。図1、図2、図3はその一例で、それぞれボストン美術館、江戸東京博物館、ヴィクトリア&アルバート博物館のデジタルアーカイブにおける

作品の情報である。図1、図2、図3の赤枠で囲った部分については後述する。浮世絵が1,000点以上所蔵されているデータベースは我々が知る限りでも20以上あり、例として、ボストン美術館で約30,000点、大英博物館で約9,000点、立命館大学アトリサーチセンター(立命館ARC)で約12,000点の浮世絵が所蔵されている。これらのうち海外のデータベースでは主に英語で、日

本のデータベースでは主に日本語でメタデータ が記述されている。



Under the Wave off Kanagawa (Kanagawa-oki nami-ura), also known as the Great Wave, from the series Thirty-six Views of Mount Fuji (Fugaku sanjūrokkei)

「富嶽三十六景 神奈川沖浪裏」

Japanese: Edo period; Woodblock print (nishiki-e);
 Artist Katsushika Hokusai, Japanese, 1760–1849
 Publisher Hishimuraya Yohachi (Eijudō), Japanese

DIMENSIONS
 Horizontal ōban; 25.8 x 38 cm (10 3/16 x 14 15/16 in.)

TYPE
 Ukiyo-e print, fūkei-ga (landscape/cityscape)

MEDIUM OR TECHNIQUE
 Woodblock print (nishiki-e); ink and color on paper

CATALOGUE RAISONNÉ
 Ukiyo-e shūka 16 (1981), p. 229, horizontal ōban #18.21

CLASSIFICATION
 Prints

ACCESSION NUMBER

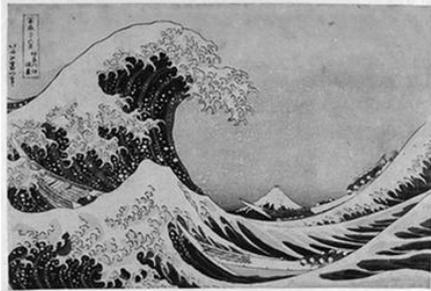
図 1：ボストン美術館の作品情報例



東京国立近代美術館蔵

番号	92202743
作品名/資料名	富嶽三十六景 神奈川沖浪裏
作家/制作者	葛飾北斎
制作年	天保2-5年(1831-1835)頃
所蔵館	江戸東京博物館
寸法	26.1* 38.5*

図 2：江戸東京博物館の作品情報例



The Great Wave; Thirty-six Views of Mount Fuji; In the Hollow of a Wave off the Coast at Kanagawa; Thirty-Six Views of Mount Fuji

Object:	Woodblock print
Place of origin:	Japan (made)
Date:	ca. 1830-31 (made)
Artist/Maker:	Katsushika Hokusai, born 1760 - died 1849 (artist)
Materials and Techniques:	Colour print from wood blocks, on paper
Credit Line:	Gift of the Misses Alexander
Museum number:	E.4823-1916
Gallery location:	Prints & Drawings Study Room, level F, case TOPIC, shelf JP

図 3：ヴィクトリア&アルバート博物館の作品情報例

我々はこれまでに、国内外に散在する複数の浮世絵データベースに対し多言語横断検索を実現する、浮世絵画像横断検索システム FeSSU[1]を構築してきた。しかし、このシステムでは、同じ作品でも所蔵されているデータベースが異なれば、異なる浮世絵データとして認識され、検索結果が煩雑になってしまうという問題点がある。

浮世絵の研究者にとって、ある組織が所有している浮世絵が本物かどうかを知るために、類似作品を見つける必要がある。また、ある浮世絵がいつ頃、どのような経緯でどの組織に渡ったのかを調べるために、どの組織がその浮世絵、あるいはその同じ版の絵、刷り方の異なる同じ絵を持っているかを知ることが重要である。このような複数の情報源に跨って散在する作品を見つけるために、テキスト検索やメタデータ検索を用いることが考えられる。しかしながら、テキスト検索やメタデータ検索を用いて類似作品を見つけるには、次の問題が挙げられる。まず、テキスト情報がない、または情報が少ないという問題がある。これは表 1 に示すようにデータベースによってメタデータ項目の数に差があることを示す。表 1 のメタデータ項目数については、図 1 の赤枠で囲った部分のように、一つの項目に複数の情報が含まれている場合でも 1 個として数えている。このように本来は複数のメタデータ項目であるものが、データベースによっては一つの項目にまとめられていることも問題としてあげられる。二つ目の問題点として多様な表記・名称・使用法があることが挙げられる。ここでの多様な表記とは"Soshu"と"Sosyu", "Katsushika"と"Katusika"のようにローマ字表記にする際に生じる違いを示し、多様な名称とは"Sagami Province" (相模州) と"Sosyu" (相州) の様に別名や略称による違いを示し、多様な使用法とは『相州江の島』と『相州～江の島～』のような文字や記号の使い方の違いを示す。

また、浮世絵の研究者からは、複数のメタデータを比較することで、データの間違いを修正し、

より正しいデータを得たい、不足している情報を補完する、あるいはより詳細なメタデータを得たいという要望もある。

そこで本稿では、異種データベース間における同一作品の同定を可能にする手法を提案する。異種データベース間における作品の比較を行う上で、問題となるのが表記法や言語の違いである。ここで表記法や言語の違いでは、表 2 や図 1~3 の赤枠部分に示すように、同一の浮世絵のメタデータでもデータベースごとに表記や言語が異なることを示す。様々なデータベースにおける表記方法を調査の結果、基本的に次の 3 種類に分類できることが分かった。原題、原題の読みをローマ字表記したもの（以後本稿では「音訳」と呼ぶ）、原題の英訳である。これらの記述方法の例として、原題で『東海道江尻田子の浦略図』となっている作品名の音訳は『Tokaido Ejiri Tago no uraryakuzu』などであり、英訳の一例として『Tago Bay near Ejiri on the Tôkaidô』がある。

そこで表 1 に、浮世絵メタデータを多く公開しているデータベース 10 件で、作品名情報の表記に原題、音訳、英訳の表記が存在するかについて、概略を調べたものを記載した。表 1 中の「○」はその表記がほとんどの場合存在する、「×」は全くその表記が存在しない、「△」はレコードによってその表記が存在する場合と存在しない場合があることを示す。表 2 に示すように、実際にデータベースによって様々な表記や言語が使用されている。

これらの表記の違いにより、作品名のメタデータをそのまま用いて他のデータベースを検索しても同一作品の情報が取得できない場合が多い。そこで、提案手法では、検索元の作品のメタデータと検索対象のデータベース群のメタデータとの類似度を算出し、類似度の高いレコードを同一作品として提示する。

表 1: データベースによる作品名の表記の違い, 及びメタデータ項目数, 浮世絵の所蔵数

	データベース	原題	音訳	英訳	項目数	浮世絵の所蔵数
国内	立命館大学アトリサーチセンター	○	○	×	61	約 12500
	国立国会図書館サーチ	○	○	×	21	約 10000
	江戸東京博物館	○	×	×	6	約 6100
海外	アメリカ議会図書館	×	○	○	16	約 2740
	大英博物館	△	○	○	15	約 9000
	ボストン美術館	○	○	○	9	約 30000
	ヴィクトリア&アルバート博物館	×	×	○	8	約 400
	メトロポリタン美術館	△	△	○	10	約 4200
	アシュモレアン博物館	×	×	○	12	約 640
	シカゴ美術館	×	△	○	8	約 2100

表 2：浮世絵データベースに散在する同一作品の例

作品名	画像	データベース
『富嶽三十六景』(原題), 『神奈川沖浪裏』(原題)		立命館アトリサーチセンター
Kanagawa-oki nami-ura (音訳)		大英博物館
Under the Wave off Kanagawa (英訳)		ボストン美術館

2. 関連研究

複数の異種浮世絵データベース間の統合的なアクセス実現のために我々がこれまでに開発してきた **Linked Data** を用いた検索システム[2]がある。このシステムでは前述の **FeSSU** の検索結果のレコードに関連する情報を検索するものである。ここでは検索結果のレコードの一つから作者情報を取得し、典拠データを通して **Linked Data** と呼ばれるデータ群を検索している。本研究では、レコードの絞り込みを行うために、[2]の検索システムのうち、典拠データを用いて作者の別名・別表記を取得する機能を用いる。この機能を用いることにより、データベース毎で異なる作者名の表記の揺れや言語の違いに依らず、レコードの絞り込みが可能になる。

浮世絵画像を対象とした検索サイトとして、**Ukiyo-e.org**[3]がある。このサイトの特徴として、浮世絵を所蔵している様々な組織が提供しているデータベースから、ある時点での全データを収集し、類似画像検索の技術を用いて同一作品を検索できる仕組みを提供している。画像の類似度を測ることで同一作品かどうかを判定しているので、別々のデータベースに散在する同一作品の画像を提示できる。

本研究の **Ukiyo-e.org** との相違点は、あらかじめ各データベースの全データを収集せず、検索を行うたびに各データベースにアクセスすることで同一作品の同定を行っている点である。このように検索するたびに処理を行うことにより、常に最新のデータベースに対して同定を行うことができるというメリットがある。

またもう一つの相違点として、画像ではなくメタデータを用いて同一作品であるかを判定している点が挙げられる。これにより、データベースで画像が公開されていないでも検索が可能であるというメリットがある。また浮世絵は版画であるので、同じ原画を別の版で刷った絵（「異版」と呼ばれる）が存在する。類似画像検索では異版同士では画像的な差がほとんどなく区別ができない。専門家から見れば、このような異版の情報は価値があり、メタデータを用いることにより異版の区別を踏まえた検索が可能になるというメリットがある。

異なるデータベース間で同一レコードの同定を行う研究は古くからあり、同じ言語同士のレコード同定であれば、主に書誌情報を対象とした同一レコードの同定に関して様々な研究が行われている[4]。これらの研究は本研究においても同言語同士の作品名の比較の際に参考となる。具体的には、例えばメタデータ項目ごとの類似度を算出する際の編集距離の拡張として **Smith-Waterman** アルゴリズムの適用を提案している[5]。これは編集距離の置換操作によるコストを文字ペアごとに定義することで大文字/小文字の違いを無視するなど柔軟な対応が可能になるものである。また適応的に距離関数を獲得する手法として、アフィンモデルで拡張した編集距離に対して、確率遷移のパラメータを **EM** アルゴリズムにより推定する方法が提案されている[6]。これにより連続した削除操作に過剰なコストがかかることが回避することが出来る。また文字列の類似度だけでは同一か否かの判定が行えない場合に、略称や異表記などの変換ルールを事例から自動学習する手法も提案されている[7]。

これらの研究は参考になる部分も大いにあるが、表 2 の例に示すような言語が異なる場合にも対処を行う点で本研究と異なる。浮世絵の作品名は歌舞伎役者の名前や地名がそのまま作品名にされていることが多い等の浮世絵固有の特徴があり、これらの特徴に対処することも必要である。

3. 提案手法

異種データベース間の同一作品の同定を実現するために、異なる二つのデータベース間のメタデータの項目の要素の類似度による比較を行う。図 4 の提案手法の流れに示すように、提案手法では作者で比較するレコードの絞り込みを行った後、作品名による比較を行っている。同定を行うたびにデータベースを参照するので、判明した同一作品などの情報は蓄積や今後の検索等に利用はしない。

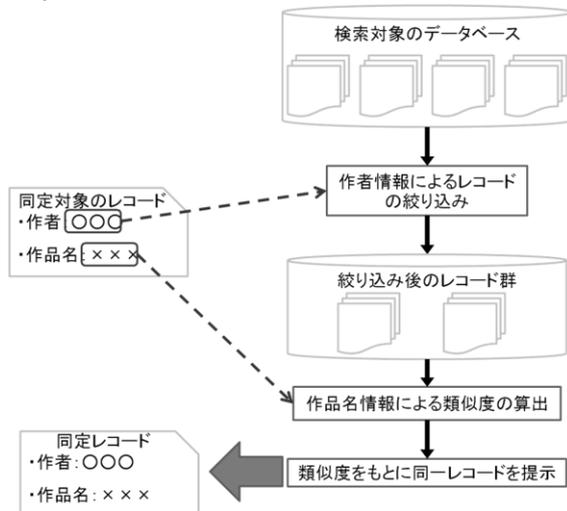


図 4：提案手法による処理の流れ

3.1 作者情報による絞り込み

提案手法の流れとしては、まずメタデータ項目のうち浮世絵の作者（絵師）の情報を用いて同定するデータの絞り込みを行う。データベースによっては作者の表記が異なる場合があり、同定を行う作品の作者名と検索対象のデータベースの作者名を比較した場合、異なる作者の作品と比較を行う可能性がある。そこで我々がこれまでに提案した複数の典拠データを検索する手法[2]を用いて作者名の表記の一覧を取得する。このシステムは様々なデータベースの典拠データが検索可能であり、検索対象のデータベースの典拠データが用いられている場合もある。その場合は検索対象のデータベースの典拠データに記されている作者名の表記を用いて絞り込みを行う。その他の場合は、複数の表記の中から、姓名両方ある表記、同定対象と同言語であるもの等、一致する可能性の高い表記をクエリとして順に作者で検索を行っている。例えば図 5 に示すように同定対象のレコードの作者名が"Hiroshige, Utagawa"であった場合、その作者名をクエリとして[2]のシステムを用いて検索を行うことにより、その作者の複数の別表記（例："Ando, Hiroshige","Hiroshige 1sei"等）を取得できる。その後、取得した別表記群のうち姓名があり、検索対象のデータベースの表記の言語と同じもの（ここでは英語とする）である"Ando Hiroshige", "Utagawa Hiroshige"等の作者名表記をクエリとする。それらのクエリを用いて順に検索対象のデータベースを作者名検索し、レコードが一つでも検出されなければ作者名による絞り込みを終了する。

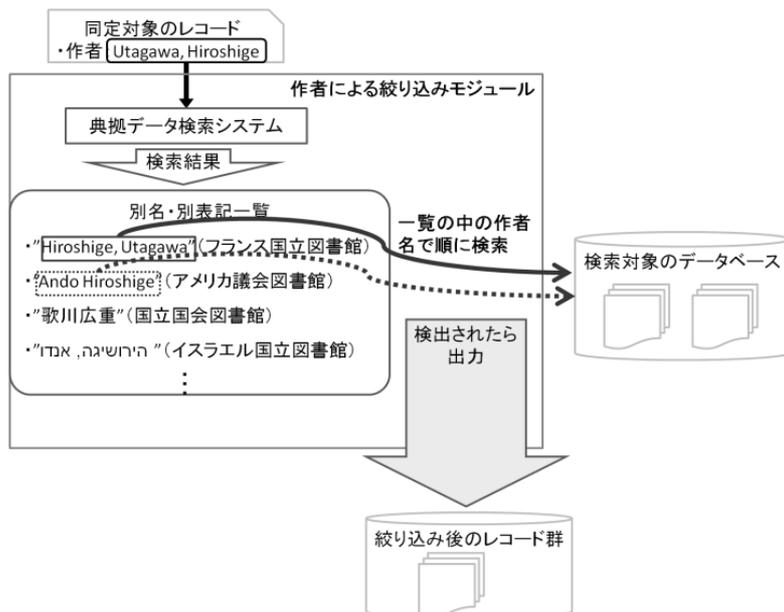


図 5：作者名による絞り込みの例

異種データベース間のメタデータを比較する際、あるデータベースの属性が他のデータベースにおいてどのメタデータ項目に当てはまるかを同定する必要があるが、本研究ではこのメタデータ項目のマッピングは人手で行う。

3.2 作品名の比較

次に、比較するデータベースの作品名情報の比較を行う。作品名情報を用いる理由としては、ほぼ全ての作品に付与されている属性値であるため比較が可能である点、作品名情報は他の属性に比べて、属性値の重複が少なく、絞り込みに有用な点が挙げられる。絵の寸法、作成年等の他のメタデータに関しては、今回は使用しない。

作品名情報を用いて比較する際、その表記が、原題、音訳、英訳の3種類の記述方法のどれに当てはまるかを特定する。今回はこの特定は手動で行った。

その後、特定した種類により、表3の対応表に示す手法を用いて類似度の計算を行う。比較するデータベースに同定対象と同じ種類の表記がある場合は、同じ種類同士で類似度を算出する。しかしデータベースによっては、作品名に関するこれら3種の情報のうち一部しかない場合もある。本研究では同言語同士（原題と音訳は同言語と見なす）の比較と異言語間の比較でそれぞれ別の比較方法を用いることで、作品名情報の記述の種類が異なる場合にも対処できるようにしている。

3.2.1 同表記同士の比較

同表記同士での比較方法として、ここでは主に編集距離を用いる。編集距離（レーベンシュタイン距離）とは、二つの文字列の類似度を示す指標であり、二つの文字列を同じ文字列にするために文字の挿入、置換、削除を行い、それらの操作の回数を距離とするものである。例として図6に、文字列「東都駿台」と「東都：駿河台」、「東都駿台」と「東京都お台場」の編集距離の算出過程を示す。この例において編集距離は「東都駿台」という文字列が「東都駿河台」と「東京都お台場」を比べてどの程度異なるかを示す基準であるので、「東京都お台場」より「東都駿河台」の方が「東都駿台」に類似していることが分かる。提案手法ではこの距離の逆数を類似度としている。編集距離の算出における挿入・置換・削除のコストは予備実験により置換の値を2、他は1に設定している。

「東都駿台」と「東都駿河台」の場合

	操作	コスト
1. 東都駿河台	「河」を削除	+1
2. 東都駿台		
		計:1
		編集距離=1

「東都駿台」と「東京都お台場」の場合

	操作	コスト
1. 東京都お台場	「京」を削除	+1
2. 東都お台場		
3. 東都駿台場	「お」を「駿」に変換	+1
4. 東都駿台	「場」を削除	+1
		計:3
		編集距離=3

図6：編集距離の例

3.2.2 異言語間（音訳と英訳）の比較

音訳と英訳を比較する場合、異言語同士であるため、編集距離を用いた比較では、類似度の算出に適していない。そこで本研究は日本語の固有名詞を英訳した場合、元の日本語を音訳したものが訳語となることが多い性質に着目し比較を行う。比較対象の二つの作品名に存在する単語間で固有名詞の単語だけで比較を行い、一致する単語が多いほど類似度を高くする。比較する単語が日本語の固有名詞か否かの判別は、英和辞書にその単語が記載されているか否かで行っている。英和辞書はEDR電子化辞書¹を用いている。しかし作品名によっては、たとえば“Garden Chrysanthemum”（原題は「庭前の菊」）のように固有名詞が一つも出現しない場合がある。そこで同定を行うレコードの作品名に固有名詞が一つも出現しない場合、英訳の作品名における各単語を英和辞書を用いて翻訳し、その読み情報について比較を行う。

異言語同士のレコードを比較する際、レコードにない情報を使わなければ単語が一致しない場合がある。例えば、あるデータベースでは“Sagami Province”（相模州）と表記されているのに対し、別のデータベースでは“Soshu”（相州）と略されている場合がある。そこで今回は、日本の旧国名とその略号の情報を用いて変換を行う。

¹http://www2.nict.go.jp/out-promotion/techtransfer/EDR/J_index.html

原題と英訳を比較する場合は、原題を音訳化し、音訳と英訳の比較方法と同様の処理方法で類似度を算出する。原題を音訳化する方法は、原題に対して形態素解析器のMeCabを用いて原題全体の読み情報を取得し、その読み情報をへボン式のローマ字に変換するものとしている。へボン式の表記にする理由としては欧米でよく利用されている形式であることが挙げられる。最後に算出された類似度をもとに類似する作品を提示する。

図7は、比較する二つのデータベースの記述の種類が音訳と原題の場合の処理の例である。ここでは同定対象の作品の作者が"Katsushika, Hokusai"であるので、"Katsushika, Hokusai"をクエリとして比較対象のデータベースを検索する(1)。その後検出されたレコードに対して、同定対象の作品名である"Tokaido Ejiri Tago no ura ryaku-zu"との類似度をそれぞれ算出し、類似度の高い「東海道江尻田子の浦略図」をユーザに提示している(3)。

表3: 表記の種類毎の処理の対応表

	原題	音訳	英訳
原題	編集距離	読みをローマ字化→編集距離	原題を音訳→固有名詞のみ比較
音訳		編集距離	固有名詞のみ比較
英訳			編集距離

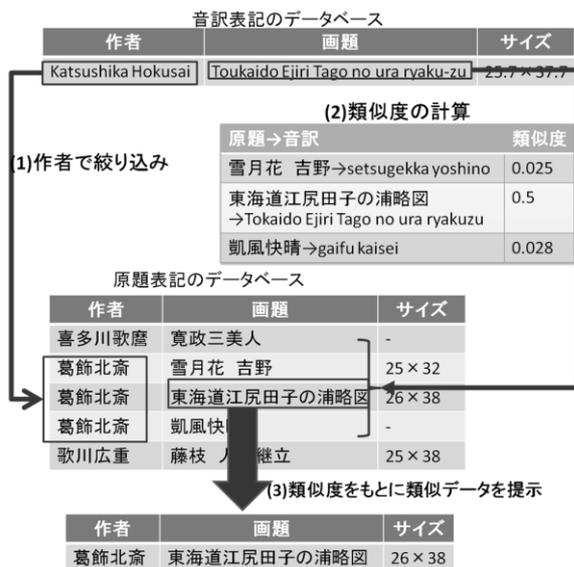


図7: 提案手法による処理の例

4. 評価実験

提案手法による同定精度の評価を行った。表3に示す表記の種類毎の処理に対して、それぞれ評価を行った。使用するデータベースは同定対象と

してボストン美術館、比較対象として大英博物館を使用し、計38件の作品に対して同定を行った。正解レコード(同一作品)は、作者による絞り込みを行ったレコード群に対して人手で作成した。従って、本来正解であるレコードに作者情報がない、または表記ミスがあった場合、正解レコードが得られない。また人手による作成の際、正解レコードか否かの判断が出来ない場合は全て不正解として扱っている。評価指標としてMAP (Mean Average Precision)を用いる。MAPは、各問合せ、すなわち1作品ごとに平均適合率をもとめ、全ての問合せの平均適合率の平均をとったものである。平均適合率の算出方法としては、まず比較対象のレコード群を類似度順に並べ、上位から順に判定範囲を増やしていく。判定範囲内の再現率が変化した時点で、判定範囲内の適合率をもとめ、判定範囲の再現率が1になったら時点で判定範囲の増加を終了する。最後にそれまで求めた適合率の平均をとる。音訳と英訳の組み合わせごとの実験結果を表4に示す。

また旧国名とその略号による変換、英和辞書使用の影響を調べるために、表5に辞書・情報の使用・不使用の場合のMAPを記載している。

表4: 同言語と異言語の比較結果

	音訳	英訳
音訳	0.852	0.697

表5: 英和辞書・旧国名情報の使用の有無によるMAPの違い

英和辞書の使用	旧国名情報の使用	MAP
無	無	0.598
有	無	0.697
無	有	0.584
有	有	0.683

5. 考察

音訳同士の比較においてはouをōと表記する等の細かい違いはあるが、基本的に単語の綴りや単語の並びは同じである場合が多いため、正解レコードの類似度が高くなったと考えられる。同言語の比較において精度が低下した原因として、たとえば比較するレコードの作品名が“Hama-cho yori utsusu Ryogoku taika”で、他のデータベースに存在する同一作品レコードの作品名が“Meiji Juyonnen ichigatsu ni juroku-nichi shukka Hama-cho yori utsushita Ryogokutaika”のような例が挙げられる。この場合は片方の作品名に付加的な情報が作品名の前半にあるため、そのまま編集距離を用いると、付加情報の部分により編集距離が大幅に増加してしまう。

また、たとえば英訳が“Pines and waves at dragon cavern”(原題:「龍洞松濤」,音訳:“Ryudo Shoto”)のように、英訳を翻訳・音訳化しても一致しないものがあることも精度が下がる原因として挙げられる。

表5より、旧国名とその略号を用いた場合、精度が下がっていることが分かる。精度が下がった原因としては、旧国名情報を用いて文字列を変換することにより単語は一致するようになったが、不正解のレコードの作品名の単語と一致してしまい精度が下がったと考えられる。地名辞書による変換の一例として、“Sagami Province”を“Soshu”と変換し比較を行うことにより、10個のレコードの類似度が上がったが、そのうち正解は2個であるため適合率が低下してしまった。また“Bizen Province”を“Bishu”と変換し比較することにより、本来不正解のレコードの作品名である“Bishu Nagoya shinkei”の類似度が上がり、正解のレコードの作品名である“Bizen Tatsukuchiyama”の類似度が下がってしまうという例も見られた。一部平均適合率が向上した場合もあるが、この例のような場合が比較的多いため、全体の精度が低下してしまう。

表5より、異言語同士の比較の際、英訳の作品名の単語を辞書を用いて翻訳・音訳化し比較することにより精度が向上していることが分かる。これは上記の地名情報を用いた場合と異なり、不正解のレコードの作品名の単語に一致する確率が低いためであると考えられる。

辞書や領域固有の情報を用いることにより網羅性は概ね向上するが、適合率としては値が下がってしまうため精度が低くなることがわかる。網羅性を保持しつつ適合率を向上させる手法を検討する必要がある。

6. おわりに

本論文では、異種データベース間における、メタデータを用いた同一作品の同定手法を提案した。この手法を用いることにより、レコードの表記の違いや使用されている言語に依らず、複数のデータベースに散在する同一作品の同定がある程度可能となった。これにより、研究者にとって研究の効率化や新たな知識獲得の手助けとなることが期待できる。今回は浮世絵に限定して同定手法の提案を行ったが、浮世絵以外のデータへの適用も期待される。

今後の課題としては、文字列比較の際の固有名詞のより正確な自動判定手法の検討、比較するメタデータ項目の自動判別、同定の補助的要素として作成年や絵の寸法等の利用が挙げられる。

謝辞

本研究の一部は、文部科学省私立大学戦略的研究基盤形成支援事業「芸術・文化分野の資料デジタル化と活用を軸とした研究資源共有化研究」、

文部科学省科学研究費補助金基盤研究(C)「多言語デジタルアーカイブの統合検索に関する研究」(研究代表者:前田亮, 課題番号:24500300)の支援を受けている。

参考文献

- 1) B.Batjargal, F.Kimura, and A.Maeda. Metadata-related Challenges for Realizing Federated Searching System for Japanese Humanities Databases. Proc. 11th International Conference on Dublin Core and Metadata Applications (DC-2011), pp. 80-85, The Hague, Netherlands (2011).
- 2) 久山岳夫, Biligsaikhan Batjargal, 木村文則, 前田亮: 動的リンク生成による浮世絵データベース間の多言語統合アクセス手法の提案, 人文科学とコンピュータシンポジウム(じんもんこん 2012) 論文集, pp.231-238 (2012).
- 3) John Resig: Ukiyo-e.org, 入手先<<http://ukiyo-e.org/>>(参照 2013-10-11).
- 4) 相澤彰子, 大山敬三, 高須淳宏, 安達淳: レコード同定問題に関する研究の課題と現状, 電子情報通信学会論文誌, DI, Vol.J88-DI, No.3, pp.576-589 (2005).
- 5) A.E.Monge and C.P.Elkan, An efficient domain-independent algorithm for detecting approximately duplicate database records. Proc. ACM-SIGMOD Workshop on Research Issues on Knowledge Discovery and Data Mining, pp.23-29 (1997).
- 6) M.Bilenko and R.J.Mooney. Adaptive duplicate detection using learnable string similarity measures. Proc. 9th ACM International Conference on Knowledge Discovery and Data Mining (KDD2003), pp.39-48, (2003).
- 7) S.Tejada, C.A.Knoblock, and S.Minton. Learning domain-independent string transformation weights for high accuracy object identification. Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2002), pp.350-359, (2002).