

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 9 日現在

機関番号：32612

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500293

研究課題名(和文) FRBR に依拠した著作および表現形の同定と同定結果を用いた検索システムの構築

研究課題名(英文) Identification of works and expressions in line with FRBR and the development of retrieval systems to utilize identified works

研究代表者

谷口 祥一 (Taniguchi, Shoichi)

慶應義塾大学・文学部・教授

研究者番号：50207180

交付決定額(研究期間全体)：(直接経費) 3,800,000 円、(間接経費) 1,140,000 円

研究成果の概要(和文)： 現行の図書館目録のメタデータ(書誌レコード、典拠レコード)を、FRBR(「書誌レコードの機能要件」)概念モデルに準拠したものとなるよう、著作の効率的な同定法かつわが国固有の書誌・典拠レコードに適合した同定法を提案し、その有効性を検証した。古典著作、翻訳資料などの資料タイプごとに、必要な同定基準を検討し、実際に同定作業を行い、同定結果を公開した。並行して、人手による同定作業を補完する目的で、機械的な同定法を組み合わせることの有効性を検証した。また、それら同定結果を有効に用いるためのFRBR型の検索システムを構築した。他方、表現形については同定に大きな困難が伴うことを確認した。

研究成果の概要(英文)： To convert existing Japanese bibliographic and authority records in accordance with the Functional Requirements for Bibliographic Records (FRBR) conceptual model, methods of identifying works and expressions included in resources have been proposed and their usefulness and feasibility have been validated. Judgment criteria for identifying works were built for some categories such as Japanese classical works and musical works, and works were manually identified using those criteria. Methods of identifying works automatically were also investigated, and the results were combined with the manually identified work sets. Additionally, an FRBRized OPAC was developed to utilize effectively the identification results. In contrast, expression identification was abandoned because of its difficulties.

研究分野：総合領域

科研費の分科・細目：情報学・図書館情報学・人文社会情報学

キーワード：メタデータ 図書館目録 FRBR

### 1. 研究開始当初の背景

(1)FRBR(「書誌レコードの機能要件」)にかかわる研究として、主に概念モデルとしてのFRBRの再検証を行う研究と、既存メタデータ(書誌レコード、典拠レコード)をFRBRに依拠したメタデータへと変換し、その変換結果を効果的に活用する検索システムを構築する研究などが見られた。

(2) に関しては、FRBRの概念モデルの検証や、資料種別などを限定したときの一部修正を加えたモデルの検討などが以前から行われており、相当程度の成果と議論の蓄積がなされていた。一方、FRBRに依拠しつつ2010年に制定されたRDA(Resource Description and Access)にも、その具体的な採用に向けた検討が開始され、それに伴って概念レベルからRDAを再検討する先駆的な試みが現れつつあった。

(3)他方については、欧米では既存メタデータをFRBRに依拠したメタデータへと変換する複数のプロジェクトやシステムが当時既に存在し、FRBR型検索システムの有効性が概ね認められていた。OCLCによるWorldCatやFictionFinder、オーストラリア国立図書館によるプロトタイプ検索システム、インディアナ大学による音楽著作に限定したScherzoなどが、これに該当した。ただし、これらの先行事例においては、既存レコードに対する著作の機械的同定が適用されており、その同定性能は十分とはいえない状況にあった。これは、既存の書誌・典拠レコードがFRBRに依拠した構成や記載内容とされていない点に起因する。特に「著作」(知的・芸術的内容のまとまり)と「表現形」(著作を文字・記号列等で表現したもの)について適切に記録または指示されている範囲は限定されており、それらを越えた広範な機械的同定には限界があるためである。一方、わが国で作成され蓄積されている書誌・典拠レコードは、著作と表現形の記録において欧米のレコードと比べてさらに貧弱といわざるをえず、同定作業の困難度がさらに高い状況にあった。特に古典著作などについては機械的な著作同定は困難を極める。そこで、わが国の状況に即して、既存書誌レコードに対する主に人手による著作同定作業と照合キー方式などを用いた機械的同定を効果的に組み合わせることで同定を図ること、および書誌同定(重複書誌同定)を著作同定と組み合わせることなどが有効と考えられた。

(4)加えて、著作および表現形の同定を行うための基準(同一著作・表現形とみなす判定基準)自体がわが国の場合明確ではなく、古典著作、音楽著作などの著作の種別や表現形の種別に合致した基準が存在していない状況にあった。これら最も基本的な部分からFRBRに依拠して整備する必要があった。

### 2. 研究の目的

(1)著作および表現形の同定基準の設定と公開:FRBRにおける著作および表現形の定義に依拠しつつ、実作業に適用できるレベルの同定基準を新たに設定する。また、これら同定基準は著作の種別、表現形の種別に合致したものとす。FRBRはこのような同定基準まで踏み込んでおらず、既存の日本目録規則などにおいても該当する同定基準は示されていないため、新たな設定が必要となる。最終的には著作および表現形の同定基準を公開し、関係者による評価を受ける。

(2)効率的な著作同定法、表現形同定法の検証:わが国の書誌・典拠レコードの特性を踏まえて、人手による同定作業と機械的同定(照合キー方式など)を効果的に組み合わせることで同定を図る方式を検討し、その有効性を検証する。併せて、機械学習の効果的な適用を検討し、多数存在する機械学習方式のうち、この課題に対していかなる方式が適しているのかを検証する。また、同一資料に対して複数かつ多様な書誌レコードが作成されている状況に対応するため、機械的な書誌同定(重複書誌同定)を併せて採用する方式とする。

(3)同定結果、同定支援ツールの公開:同定作業結果をデータベースに蓄積し、同定結果から機械的に生成した著作レコードや表現形レコードを公開する。同定作業をはじめとする協力者向けの作業結果公開とWeb APIによる作業結果公開という2つの方式での公開を図る。前者は協力者向けの作業結果公開であり、分散型での同定作業を支援するツールとしての機能も備えたものとする。

(4)FRBR型検索システムの構築:著作・表現形の同定結果を含め、整備したレコード群を活用した有効なユーザ向け検索システムを開発する。FRBRに依拠して著作、表現形、表現形の間を自由に行き来できるナビゲーションを実装したシステムとする。併せて、適切なシステム評価手法の検討を踏まえて、新規開発システムと既存のFRBR型検索システム(欧米のシステム)とを対象にしたユーザによる評価実験を行う。

### 3. 研究の方法

(1)FRBRにおける著作および表現形の定義に依拠しつつ、実作業に適用できるレベルの同定基準を、著作の種別(文字を主とするテキスト系著作、音楽著作などの種別)や表現形の種別に分けて検討を行った。わが国の標準的なレコードである国立国会図書館作成の書誌レコードを主な検討材料に用い、書誌レコードが表す資料本体も一部参照しつつ検討を進めた。また、欧米等において既に公開されている基準等とわが国固有の状況等

を勘案し検討を進めた。最終的には、同定基準の適用の容易さと同定基準を適用して得られる同定結果の網羅性とのバランスを重視した同定基準を意図した。なお、同定基準の設定は FRBR 研究会との連携をもって進めた。

(2)先に設定した著作の同定基準を用いて、国立国会図書館作成の書誌レコードに対して人手による著作の同定作業を実施した。同定作業の過程で新たに生じた検討課題は同定基準の修正というフィードバックによって解決した。これら同定作業は FRBR 研究会と共同して行い、並行して同定作業を支援するためのツールの検討と開発を進めた。他方、著作の照合用のキーを生成し同定を図る方式など機械的同定法を、複数試みた。国立国会図書館作成の書誌レコードに対する実験および多数の図書館によるレコードを集めた総合目録ネットワーク（ゆにかねっと）の書誌レコードを実験対象とした。単一書誌レコードに対して、主たる著作を一つ同定する方式と、注記記載のものを含め、複数の著作を同定する方式を実験した。また、著作の機械的同定に際して、機械的な書誌同定（重複書誌同定）を併せて実行することが有効であるのか検証した。さらには、複数の機械学習方式を、著作の同定処理および書誌同定（重複書誌同定）に適用し、その性能を評価した。

(3)先の同定作業結果をデータベースに蓄積し、公開用の著作レコードを機械的に作成し、対応する書誌レコードの識別番号（ISBN、MARC レコード番号等）とともに公開した。また、先行する OCLC による Web API を参考にしつつ、同一著作として同定された書誌レコード群の識別子を Web API を用いて公開した。

(4)著作の同定結果を含め、整備したレコード群を活用するユーザ向け検索システムを開発した。システムは、著作と体现形の間を自由に行き来できるナビゲーションを実装したインタフェースとした。システム評価手法を併せて検討し、新規開発システムと既存の FRBR 型検索システムとを対象にしたユーザによる評価実験を計画した。併せて、より有効な検索システムとなるよう主題アクセス機能の向上を検討した。

(5)欧米において、FRBR に依拠し新たに策定されたメタデータ作成基準（目録規則）である RDA に対して、著作・表現形にかかわる規定を含めて、主に理論的側面からの検討を行った。当初の研究目的には含めていなかった事項であるが、今後は RDA に基づき作成・提供されるメタデータが、特に海外では増加していくものと予測されるため、RDA の理解と検討が不可欠な位置づけとなると

考え取り組んだ。

#### 4. 研究成果

(1)日本の主要な古典著作に限定し、その効率的な同定法（同一著作に対する書誌レコード群の同定）を検証するため、人手による同定作業を実施した。同定結果（判定結果）は、それぞれの書誌レコード内に同定箇所の指示を含めて記録した。実際の同定作業とのフィードバックを繰り返した結果、FRBR における著作の定義に依拠しつつ実作業に適用できるレベルの同定基準、かつわが国固有の書誌・典拠レコードに適合した基準を策定できた。こうした成果の公開は、他研究者による同種の研究を誘発する、もしくは多様な活用に結びつく可能性があり、重要な意義をもつ。また、表現形の同定や著作間の関連の同定についても、その実施可能性を検討したが、適用の困難度が高く、今後の課題とした。特に、表現形に対する同定基準および同定作業は欧米においても殆ど未着手であり実績がなく、いかなる種別に適用できるのか、適用する際にどのような基準とすべきかから検討したが、結論を得ることはできなかった。

(2)日本語に翻訳された海外著作に限定し、その効率的な同定法（同一著作の各種翻訳資料に対する書誌レコード群の同定）を検証するため、人手による同定作業を実施した。同定作業に並行して、FRBR における著作の定義に依拠しつつ実作業に適用できるレベルの同定基準、かつわが国固有の書誌・典拠レコードに適合した基準を策定した。既存の複数の翻訳資料リストを参照し、それらの間に同定結果の揺れ（すなわち同定基準の揺れ）があることを併せて確認した。なお、上記著作にかかわる表現形の同定や著作間の関連の同定についても、その実施可能性を検討したが、適用の困難度が高く、今後の課題とした。

(3)計算機プログラムによる機械的な著作同定、および同一体现形とすべき書誌レコード群の書誌同定を行った。実験対象の書誌レコードは、国立国会図書館作成によるものに加えて、総合目録のレコードを実験用に借用し使用した。古典著作等、機械的な同定が困難なものを除いては、概ね良好な著作同定結果および書誌同定結果が得られることを確認した。著作の照合用キーを生成し同定を図る方式など機械的同定法を適用し、意義ある実験結果が得られたものと評価する。ただし、対象とする書誌レコードの特性から、単一書誌レコードに対して主たる著作を一つ同定する方式に限定している。併せて、これまでに各種機関で作成されたリスト（翻訳資料リストなど）を組み入れることによって同定性能が向上することを確認した。また、複数の機械学習方式を上記の著作同定に適用したが、大規模なレコード群に対する適用の困難さがあり、また実用に耐える性能値も得られ

なかった。

(4)同定された古典著作の書誌レコード群から、著作レコードを機械的に生成した。それらレコードには、同定されたすべてのタイトルの異表記、著者名の異表記、付与されている分類記号や件名標目などが、書誌レコード番号や ISBN などの識別子に加えて記録されたものとした。これらを試作した FRBR 型検索システムとその Web API を介して公開した。対象とした著作数など未だ量的に十分とはいえないが、実際の同定結果を公開したことも含めて、十分に意義ある成果が得られたものと評価する。

(5)著作と表現形の間を自由に行き来できるナビゲーションを実装した FRBR 型検索システムを試作した。検索対象データは、先に同定処理を行った古典著作とそれらに対応する書誌レコードのみとした。既存の FRBR 型検索システムとの性能評価については、蓄積されたレコード数が少なすぎる点および性能評価の尺度等が適切に設定できなかったため、実施していない。関連して、FRBR 型の検索システムにおける主題検索機能の高度化を目指し、わが国の主題表現ツールである BSH (基本件名標目表) NDLISH (国立国会図書館件名標目表) NDC (日本十進分類法) の統合的活用の方策を検討しその有効性を検証した。特に統制語と分類項目との相互マッピングおよびそれらの階層構造における対応関係を踏まえて、相互の相違の状況を確認し、併せて統合的活用法を提案した。ただし、こうした機能は FRBR 型の検索システムに特有のものとはならなかった。

(6)FRBR に依拠し新たに策定された RDA において、著作および表現形がどのように扱われているのか、すなわち旧来の AACR2 (英米目録規則第 2 版) からどのように変更され、それによってわが国の著作・表現形の同定処理が変更を来すのかについて明らかにした。さらには、その前提となる RDA 全体の構造的把握と検討を行い、RDA は FRBR に依拠しつつも独自の概念モデルを有すると捉えるべきこと、著作・表現形を含めて集合的実体と構成的実体の扱いについては、モデルとそれに基づく記述規則の双方において未確定な部分が残されていること、RDA モデルにおいて利用者タスクと属性・関連とのマッピングおよびタスク達成における重要度評価を FRBR/FRAD に示せること、RDA 全体を DC アプリケーション・プロファイルの観点から捉えることがその理解にとって有効であることなどを明らかにした。

## 5. 主な発表論文等

〔雑誌論文〕(計 11 件)

谷口祥一, BSH/NDLISH と NDC における階層構造はどの程度一致するのか. 日本図書

館情報学会誌. 査読有. Vol.60, No.1, 2014, (印刷中)

谷口祥一, RDA (Resource Description and Access) でできることできないこと: RDA の理解に向けて. 情報管理. 査読無. Vol.56, No.11, 2014, p.758-765.

DOI:10.1241/johokanri.56.758

Shoichi Taniguchi, User tasks in RDA-based model. Cataloging & Classification Quarterly. 査読有. Vol.51, Issue 7, 2013, p.788-815.

DOI:10.1080/01639374.2013.800623

Shoichi Taniguchi, Event-aware FRBR and FRAD models: Are they useful? Journal of Documentation. 査読有. Vol.69, No.3, 2013, p.452-472.

DOI:10.1108/JD-01-2012-0004

Shoichi Taniguchi, Duplicate bibliographic record detection with an OCR-converted source of information. Journal of Information Science. 査読有. Vol.39, No.2, 2013, p.153-168.

DOI:10.1177/0165551512459923

Shoichi Taniguchi, Understanding RDA as a DC Application Profile. Cataloging & Classification Quarterly. 査読有. Vol.51, Issue 6, 2013, p.601-623.

DOI:10.1080/01639374.2013.777865

Shoichi Taniguchi, Aggregate and Component Entities in RDA: Model and Description. Cataloging & Classification Quarterly. 査読有. Vol.51, Issue 5, 2013, p.580-599.

DOI:10.1080/01639374.2013.763316

谷口祥一, RDA をモデルと記述に分けて捉える: RDA 理解のための枠組み. 図書館界. 査読有. Vol.64, No.6, 2013, p.378-390.

Shoichi Taniguchi, Viewing RDA from FRBR and FRAD: Does RDA represent a different conceptual model? Cataloging & Classification Quarterly. 査読有. Vol.50, Issue 8, 2012, p.929-943.

DOI:10.1080/01639374.2012.712631

Takuya Tokita, Maiko Kimura, Yosuke Miyata, Yukio Yokoyama, Shoichi Taniguchi, and Shuichi Ueda. Identifying Works of Japanese Classics for Construction of FRBRized OPACs.

Cataloging & Classification Quarterly. 査読有. Vol.50, Issue 5-7, 2012, p.670-687.

DOI:10.1080/01639374.2012.681605

谷口祥一, 総合目録データに対する機械的書誌同定と著作同定の試み: ゆにかねっとレコードによる実験. 日本図書館情報学会誌. 査読有. Vol.57, No.4, 2011, p.124-140.

〔学会発表〕(計 9 件)

谷口祥一, BSH と NDLISH の統合的活用の可能性: 両者の相違の積極的活用法. 2013 年度三田図書館・情報学会研究大会, 2013 年 11

月 09 日, 慶應義塾大学 (東京)

谷口祥一. BSH/NDLSH と NDC における階層構造はどの程度一致するのか: 件名と分類の統合的活用法に向けて. 第 61 回日本図書館情報学会研究大会, 2013 年 10 月 12 日 ~ 10 月 13 日, 東京大学 (東京)

谷口祥一. RDA における利用者タスク: RDA モデルの確立を目指して. 2013 年日本図書館情報学会春季研究集会, 2013 年 05 月 25 日, 筑波大学 (つくば市)

谷口祥一. RDA における集合的実体と構成的実体の扱い: モデルと記述の分離を踏まえて. 第 60 回日本図書館情報学会研究大会, 2012 年 11 月 17 日 ~ 11 月 18 日, 九州大学 (福岡市)

谷口祥一. RDA のモデルと記述: DC アプリケーションプロファイルの観点から捉える. 2012 年度三田図書館・情報学会研究大会, 2012 年 10 月 06 日, 慶應義塾大学 (東京)

谷口祥一. RDA を FRBR/FRAD から見る: RDA は新たなモデルか? 2012 年日本図書館情報学会春季研究集会, 2012 年 05 月 12 日, 三重大学 (津市)

谷口祥一. イベント概念を導入した FRBR/FRAD モデルの検討. 第 59 回日本図書館情報学会研究大会, 2011 年 11 月 12 日, 日本大学文理学部 (東京)

谷口祥一. 総合目録データに対する機械的書誌同定と著作同定の試み: ゆにかねっとレコードによる実験. 2011 年度日本図書館情報学会春季研究集会, 2011 年 5 月 14 日, 東京学芸大学 (東京)

宮田洋輔, 上田修一, 谷口祥一, 横山幸雄, 鴫田拓哉, 向當麻衣子. FRBR における「著作」実体としての日本の古典著作: FRBR 研究会の取り組み II. 2011 年度日本図書館情報学会春季研究集会, 2011 年 5 月 14 日, 東京学芸大学 (東京)

〔図書〕(計 0 件)

〔その他〕

ホームページ等

<http://web.keio.jp/~taniguchi/>

## 6. 研究組織

### (1) 研究代表者

谷口 祥一 (TANIGUCHI, Shoichi)

慶應義塾大学・文学部・教授

研究者番号: 5 0 2 0 7 1 8 0

### (2) 研究分担者

( )

研究者番号: