

DBpedia を利用したマンガの書誌データからの Work の同定

カ ブンリン¹ 三原 鉄也¹ 永森 光晴² 杉本 重雄²

筑波大学大学院図書館情報メディア研究科¹

筑波大学図書館情報メディア系²

〒305-0821 茨城県つくば市春日 1-2

{benlin,mihara,nagamori,sugimoto}@slis.tsukuba.ac.jp

概要

書誌レコードの機能要件 (FRBR) で定義された実体を参照し、マンガの著作や出版形態などの情報を提供することで、ユーザーがより簡単にマンガを探索できると考えられる。しかし、ライブラリー (図書館、博物館等) が持つ他の所蔵資料の書誌データと比べてマンガの典拠データが十分に整備されていないため、マンガの書誌データにおいて work を利用することが難しい。本研究では、Wikipedia を LOD 形式にしたデータを提供する DBpedia を利用し、書誌データからマンガの FRBR 実体の同定を行った。同定にあたっては京都国際マンガミュージアムの所蔵マンガの書誌データを用いた。

キーワード

マンガ書誌データ, work グループ (Work-set), FRBR 化, Linked Open Data, DBpedia

Identification of Works of Manga from Bibliographic Data of Comic Books using DBpedia

Wenling He¹ Tetsuya Mihara¹ Mitsuharu Nagamori² Shigeo Sugimoto²

Graduate School of Library, Information and Media Studies, University of Tsukuba¹

Faculty of Library, Information and Media Science, University of Tsukuba²

1-2 Kasuga, Tsukuba, Ibaraki, 305-0821, Japan

{benlin,mihara,nagamori,sugimoto}@slis.tsukuba.ac.jp

ABSTRACT

Manga – a Japanese term meaning graphic novels and comics – has been globally accepted. In Japan, there are a huge number of monographs and/or magazines of manga. Functional Requirements of Bibliographic Records (FRBR) provides useful concepts for readers to identify entities of manga, e.g., Work of manga, Manifestation of manga, and so on. This paper shows a study to identify works of manga in a set of bibliographic records maintained by Kyoto International Manga Museum to help readers find a comic book as an instantiation of a work. It is known that the authority data is useful to identify works from the bibliographic records. However, the authority data of manga are not rich, because manga has been recognized as a sub-culture resource and not included in library collections. In this study, we used DBpedia, which is a large Linked Open Data (LOD) resource created from Wikipedia, to identify FRBR entities of manga in the bibliographic records. The result of experiment shows that using LOD resources is reasonable to find works from bibliographic records, but it also shows the accuracy and efficiency depend on the quality of the LOD resources used.

KEYWORDS

Bibliographic Data of comic books, Work sets, FRBRization, Linked Open Data, DBpedia.

1. はじめに

マンガは近年、日本を代表するポップカルチャーの一つとして広く認知されるに至った。しかし 1990 年代までは、マンガを収集の対象としたライブラリー（図書館、博物館等）は少なかった。90 年代後半から 2000 年代に入り、日本では、広島市まんが図書館（1997 年開館）を始め、京都国際マンガミュージアム（2006 年開館）のようなマンガを主要なコレクションとして収集するライブラリーが増えてきた。海外においては、アメリカのオハイオ州立大学図書館（1999 から所蔵開始）と台湾の臺北市立圖書館中崙分館（1998 から所蔵開始）等がマンガを収集するライブラリーとして挙げられる。

マンガは同一の内容が雑誌と単行本の異なる 2 つの出版形態により出版されたり、単一の物語が複数の書籍に渡って展開されたりする。そのため、ユーザーは個別の書籍ではなく、それらを集約した単位によって、マンガの認知や探索を日常的に行なっている。こうした個別の書籍を集約した実体は、書誌レコードの機能要件（FRBR）[1]では“work”（著作）と定義されている。ユーザーはひとつの work から、その work の異なる通訳版などの“expression”（表現形）に導かれ、また新書版などの異なる“manifestation”（体現形）を発見し、ライブラリーで所蔵している“item”（個別資料）に辿り着くことが出来る。

ライブラリーでは、ユーザーの資料探索の利便性をはかるため、本や雑誌等の所蔵資料に関する書誌データが提供されている。しかし、このような書誌データの殆どは manifestation と item 単位のデータであるため、work 単位での情報の提供と検索は実現しておらず、十分な利便性を提供していない。そこで本研究では、マンガの work 単位での識別を容易に実現する書誌データの構築を目的とする。

Hickey[2]は、典拠データから抽出されたタイトルと著者名を利用した work の同定の有効性を示した。しかし、ライブラリーにとってマンガは新しいコレクションであるため、マンガの典拠データは整備されていない。一方で、インターネット上にはマンガに関する情報が多く存在している。例えば、出版社や著者が公開しているマンガの公式サイト、またファンが作成した非公式の Web サイトなどが挙げられる。また、Wikipedia にも多くのマンガ作品についての情報が記述されている。

本研究では、Wikipedia のリンク情報などのデータを Linked Open Data (LOD) として公開している DBpedia 利用し、マンガの書誌データから機械的に work を同定する手法を提案する。手法の提案に当たっては、京都国際マンガミュージアムから提供された同館の所蔵マンガの書誌データに提案手法を適用し、work 同定の試行を行った。

2. マンガの work（著作）

FRBR は書誌データの利用者の関心の対象として、知的・芸術的活動の成果に関する実体の集合を第 1 グループとして階層的に定義している。work は個別の知的・芸術的創造を指し、第 1 グループの中の最上位の実体である。FRBR は同様に著作の内容を表現する知的・芸術的実現を指す実体、著作の表現形の物理的な形態を指す実体、及びその単一の例示となる実体をそれぞれ、expression, manifestation, item として定義している。また、FRBR は、work のインスタンスが幾つかの個別の work インスタンスによって構成されることが可能であると指摘している。こういった work のインスタンスは、集合的実体、または構成的実体であると定義している。

マンガは個別のエピソードが継続的に刊行され読者に提供されることが一般的である。例えば、1 冊のマンガ雑誌の一号の中に、異なる著者が描いたマンガが複数掲載され、一方で同一の著者が描いたストーリーの連続するエピソードは個別の単行本として集約され出版されている。このような雑誌や単行本は、著者と編集者が創作した集合的（構成的）実体であると考えられる。また同時に、雑誌や単行本自身も個別の work として考えられる。

Morozumi[4]はマンガのためのメタデータについての基本的な要求要件としてマンガメタデータフレームワーク (MMF) を定義し、その中でマンガに関する書誌を記述するための FRBR に基づいたモデルを提案した。MMF は書誌レコードの記述対象となるストーリーの構造を title、story、episode という 3 つの実体とその関係によって定義した。この 3 つの実体は 1 対多の関係、即ち 1 つの title は複数の story から成り、1 つの story は複数の episode から成る関係を持つ。更に MMF はストーリーの構造を示す実体をそれぞれが FRBR の work、expression、manifestation の実体を持つように定義している。

本研究では、MMF に基づいて図 1 のようにマンガの書誌データに含まれる実体と実体間の関連についてのモデルを定義した。一般的なマンガの実体として、個別のマンガ作品、あるマンガの一部となっている個別のストーリー、及び複数のマンガを集めて 1 つにまとめたものの 3 種類の実体が存在していると考えられる。本モデルではこれらの実体をそれぞれ、manga-title、story-title 及び manga-compilation と定義した。manga-title とは独立なタイトルを持ち、個別の著者または著者グループによって創作される単一の work または work の集合である。story-title とは manga-title を構成する個別のストーリーである。同じ manga-title を構成する story-title はお互いに共通するメインタイトルを持ちながら、それぞれ独立なタイトルを持っている。manga-compilation は編集によって製作される複数のマンガの集合であり、例えば、マンガ雑誌が当てはまる。また図 1 の矢印はそれぞれ FRBR の第 1 グループに基づく階層構造、及び各階層内における manga-title、story-title と manga-compilation の間の関連を示している。各階層内では is part of と is carried in という 2 つの関連が存在している。is part of は story-title が manga-title の一部であることを示す。is carried in は manga-title が manga-compilation に掲載されることを示す。更に、各階層間では FRBR の第 1 グループの実体間の関連を継承する。個別の work 実体は一つ以上の expression によって表現されることを示す is realized through、個別の expression 実体は一つまた一つ以上の manifestation によって具体化されることを示す is embodied in の関係をそれぞれ持つ。

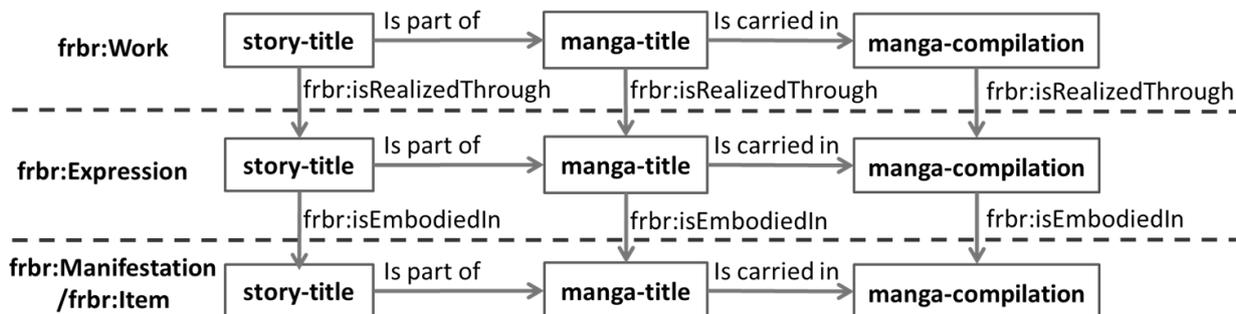


図 1. マンガ実体と実体間の関連

本研究では、本章で述べたマンガの実体及び、実体同士の関連、即ち図 1 で示した階層内の関連と階層間の関連をマンガの書誌データから識別することで work を同定することを目指す。

3. LOD リソースを利用したマンガの work の同定と DBpedia

書誌データからの work の同定と FRBR 化については、人手による同定作業のコストが高いため、機械的に同定する取り組みが行われている。Hickey の研究より、MARC のような特定の構造に従って作成された目録レコードの典拠データから抽出された著者とタイトルの組合せが work を同定するための有効なキーとして利用できるとしている。しかし、現状では 2 章のような構造に従って work を識別するのに有効なマンガの典拠データは完備されていない。そのため本研究では LOD リソースをマンガ典拠データの情報を補足する辞書として利用し、work を識別する。そのプロセスは以下になる。(図 2)

- 1) まず書誌データ内のタイトルとサブタイトルについて空白や特殊記号の削除などの正規化を行う。次に

正規化されたデータセットと著者典拠データを結合させる。最後に、(タイトル、サブタイトル、著者)と(タイトル、著者)という2つの同定キーを利用して、結合したデータセットから manga-title と story-title を同定する。この過程で作成したデータを Semi-FRBRized データセットと呼ぶ。

- 2) Semi-FRBRized データセットのリソースと LOD リソースと関連付ける。関連付けるためのプロパティとしてタイトル、サブタイトル、著者の記述を用いる。
- 3) 書誌データ、Semi-FRBRized データセットと LOD リソースの情報を整合し、適切なマンガの実体と実体間の関連を同定することで、FRBRized データセットを作成する。

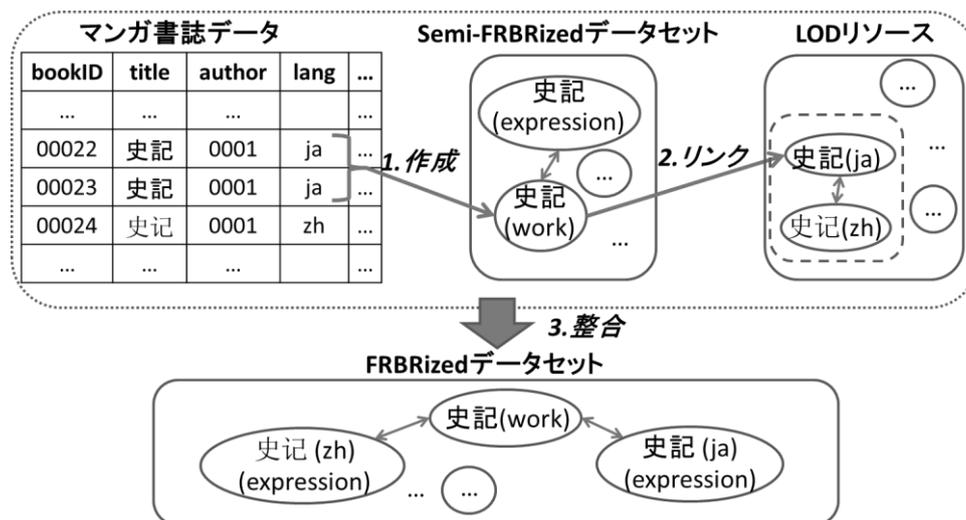


図 2. FRBRized データセットの作成プロセス

図 2 は、FRBRized データセットの作成プロセスを「史記」というマンガ作品を例にして示したものである。書誌データの中にマンガの「史記」という work についてのレコードが幾つか存在している。このレコードには日本語の「史記」とその中国語版である「史记」についての記述が含まれている。この2つの表現(expression)が同じ work であることを示すレコードは存在していないため、書誌データの情報のみを用いて work「史記」を識別するのは難しい。しかし、日本語のレコードのみで Semi-FRBRized データセットを作成し、LOD リソースと関連付けることで、この2つの expression は同じ work であることを示すもの情報をえる。その情報を参照することで、work「史記」と expression である日本語の「史記」及び中国語の「史记」という3つの実体と実体間の関連を識別することができると考えられる。

4. マンガ書誌データから work の同定の試行

4.1 試行の対象として利用したマンガ書誌データ

本研究では3章で示した work 同定のプロセスの有効性を検証するために、書誌データとして京都国際マンガミュージアム (KMM) の OPAC 用データ、LOD リソースとして DBpedia Japanese をそれぞれ利用して work の同定を試行した。

KMM の書誌データは41個のテーブルによって構成されるリレーショナルデータベースに保存されている。これらのテーブルは雑誌、単行本、著者典拠の3つのグループに関する情報として分類できる。雑誌と単行本のレコードは主に manifestation と item レベルの情報についての記述である。この試行では5000件のマンガ単行本の日本語の書誌レコード、1107件の雑誌タイトル、著者典拠データを利用した。KMM の書誌レコードではタイトル及びサブタイトルが記述されているが、レコード間の work-expression の関係を示す情報と雑誌と連載作品間の関連を示す情報も記述されておらず、manga-title と manga-compilation との is carried in の関連と work と expression との is realized through の関連を識別することができない。

この試行では、is carried in と is realized through の関係を同定するために DBpedia Japanese を利用する。DBpedia

とは Web 上では Wikipedia から情報を抽出して LOD (Linked Open Data)として公開するプロジェクトであり、その日本語版として DBpedia Japanese が公開されている。DBpedia Japanese では多くのマンガリソースが記述されている。今回の試行では、マンガのリソースとそのマンガが連載した雑誌のリソースを関連付けている「dbpprop-ja:掲載誌」と、異なる言語で書かれた同じリソースの関連を記述している「dbpedia-owl:wikiPageInterLanguageLink」を利用する。これらの情報を機械的に参照することで、is carried in と is realized through という2つの関連を同定できると考えられる。

4.2.4.3 節では KMM の書誌レコードと DBpedia Japanese を利用し、3章で示したステップでの work の同定について述べる。

4.2 Semi-FRBRized データセットの作成

ステップ 1 では、まず work レベルで記述されている DBpedia Japanese のリソースと関連付けるための work のインスタンスを含む Semi-FRBRized データセットを作成する。work は言語によって区別されない実体である。しかし、対訳に関する情報がない場合、機械的に work のインスタンスを同定することは困難である。そのため、単一言語（今回は日本語を採用した）で書かれたレコードのみを対象としてプロセスのステップ 1 を行い、work の候補となるインスタンスを作成する。また、ステップ 1 の段階では、work の候補のリソース

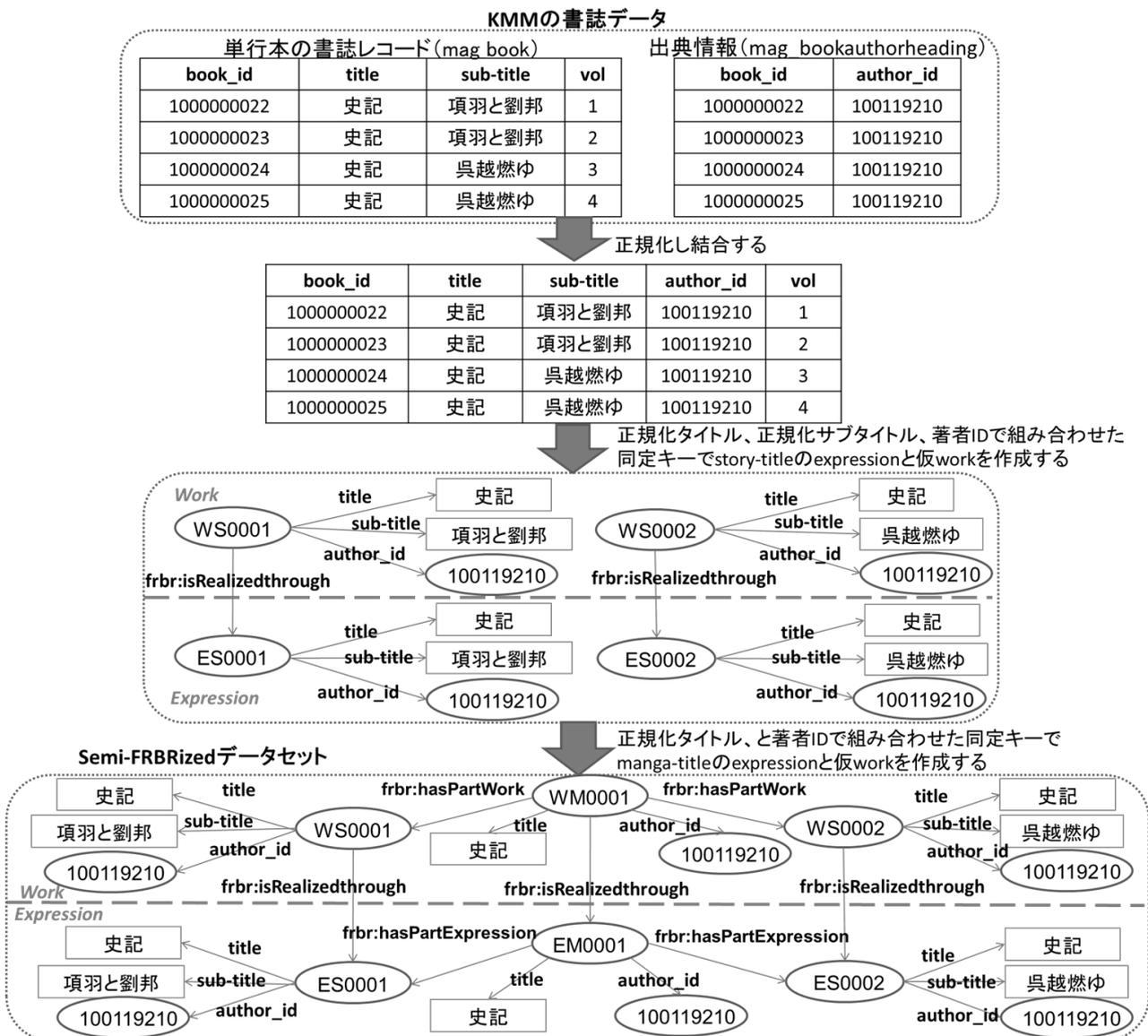


図 3. ステップ 1 の適用手順

は実 work の expression に相当するため、Semi-FRBRized データセット内の work と expression 間の関係は 1 対 1 であるのに対し、最終段階の FRBRized データセットでは 1 対 n work と expression 間の関係も含む。

図 3 はステップ 1 の適用の手順を示したものである。まず、同定キーの一部分としてタイトルとサブタイトルの標記を正規化し、出典情報を記述するテーブルを参照し、正規化されたデータセットと著者典拠を統合する。次に、“正規化タイトル、正規化サブタイトルと著者 ID” という同定キーを利用し story-title を同定する。図 3 では、expression 階層の story-title のインスタンスとして、ES001 と ES002 を作成する。また、expression のインスタンスに沿って、story-title の work の候補のインスタンスである WS001 と WS002 を作成する。そして、“正規化タイトルと著者 ID” という同定キーを利用し manga-title を同定する。図 3 では、work 候補 と expression 階層の manga-title のインスタンスとして、WM001 と EM001 を作成する。

ステップ 1 では、Semi-FRBRized データセットのリソースのうち、work 候補のインスタンスとして、5000 件の日本語の単行本書誌レコードから、2684 件の manga-title と 210 件の story-title を抽出した、また、雑誌の書誌レコードから 1107 件の manga-compilation を抽出した。

4.3 DBpedia Japanese を利用した実体間の関連の発見

ステップ 2 で DBpedia Japanese のリソースと関連付け、ステップ 3 で書誌データ、Semi-FRBRized データセットと LOD リソースの情報を整合することで、work 候補の関連付けを行い、構造を持つ集合として同定する

ステップ 2 と 3 は Semi-FRBRized データセットの work 階層のインスタンス、つまり work 候補のインスタンスを対象として行う。全ての work の候補は対応する DBpedia Japanese のリソースと関連し、「dbpprop-ja:掲載誌」と「dbpedia-owl:wikiPageInterLanguageLink」を参照し、manga-title と manga-compilation 間の関連である is carried in、及び 1 対 n の work-expression の関連である is realized through を作る。

また、Wikipedia は http://ja.dbpedia.org/resource/ONE_PIECE のように記事のラベル (タイトル) から URI を作成する。そのため、ステップ 2 では、Wikipedia と同じルールで work の候補のタイトルから URI を生成する。このうち、マンガではない、または実際の Wikipage のリソースを表現していないものを排除する。そして、条件に排除されない URI を正解とみなし、work 候補のインスタンスと rdfs:seeAlso というプロパティで関連づける。ステップ 2 では、Semi-FRBRized データセットのうち、558 件の manga-title、及び 191 件の manga-compilation の work の候補のインスタンスは、それらと対応する DBpedia Japanese のリソースと関連付けることができた。関連付けた後のデータセットの構造は図 4 に示す。

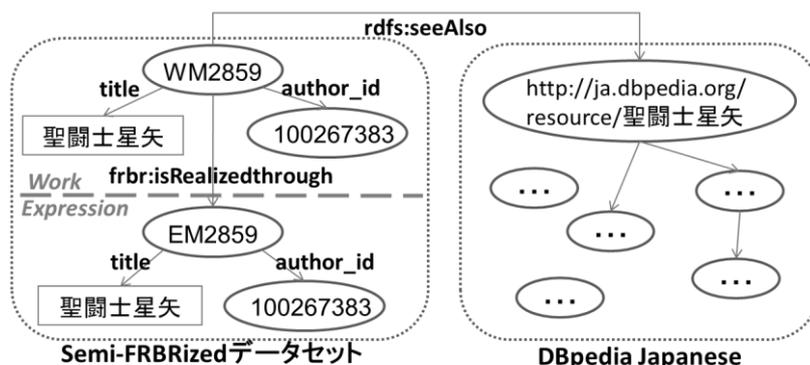


図 4. ステップ 2 の結果セットの構造

ステップ 3 では、図 5 のように「dbpprop-ja:掲載誌」の情報を直接参照することで、125 対の manga-title と manga-compilation を関連付けた。更に、「dbpedia-owl:wikiPageInterLanguageLink」の情報を間接的に参照することで、1 対 n の work-expression 関連を付けた。具体的な手順を図 6 に示す。

- i. URI 形式で表現されている「dbpedia-owl:wikiPageInterLanguageLink」の値から異なる言語のタイトルと

言語コードを取得し、`rda:parallelTitleProper` というプロパティで、`work` の候補の属性として記述する。この作業を通して、`work` の候補を言語に制限されない実 `work` に変換することができる。

- ii. **Semi-FRBRized** データセットの作成作業で除外した外国語版のマンガの書誌レコードから、取得したタイトルと言語コードと一致するレコードの群を探す。
 - iii. ステップ ii で見つかったレコードの群に対し、**Semi-FRBRized** データセットの作成作業を行い、`work` の候補を作成せずに、`expression` レベルまでの FRBR 化を行う。
 - iv. `frbr:isRealizedThrough` というプロパティで `work` とそれに対応する `expression` の関連を作る
- 上述のステップを行うことで、558 件の `manga-title` と 191 の `manga-compilation` の実 `work` を作成した。

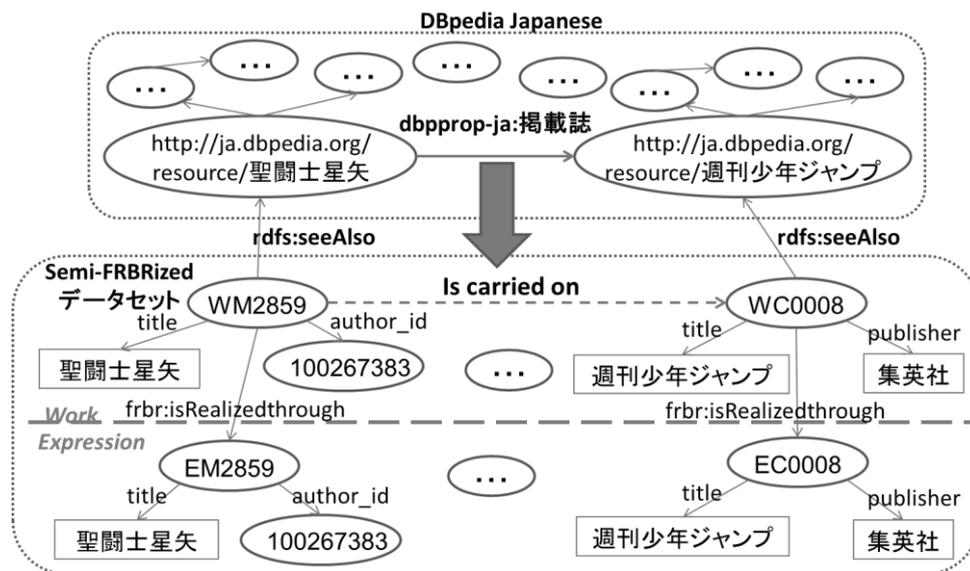


図 5. manga-title と manga-compilation の関連の作成

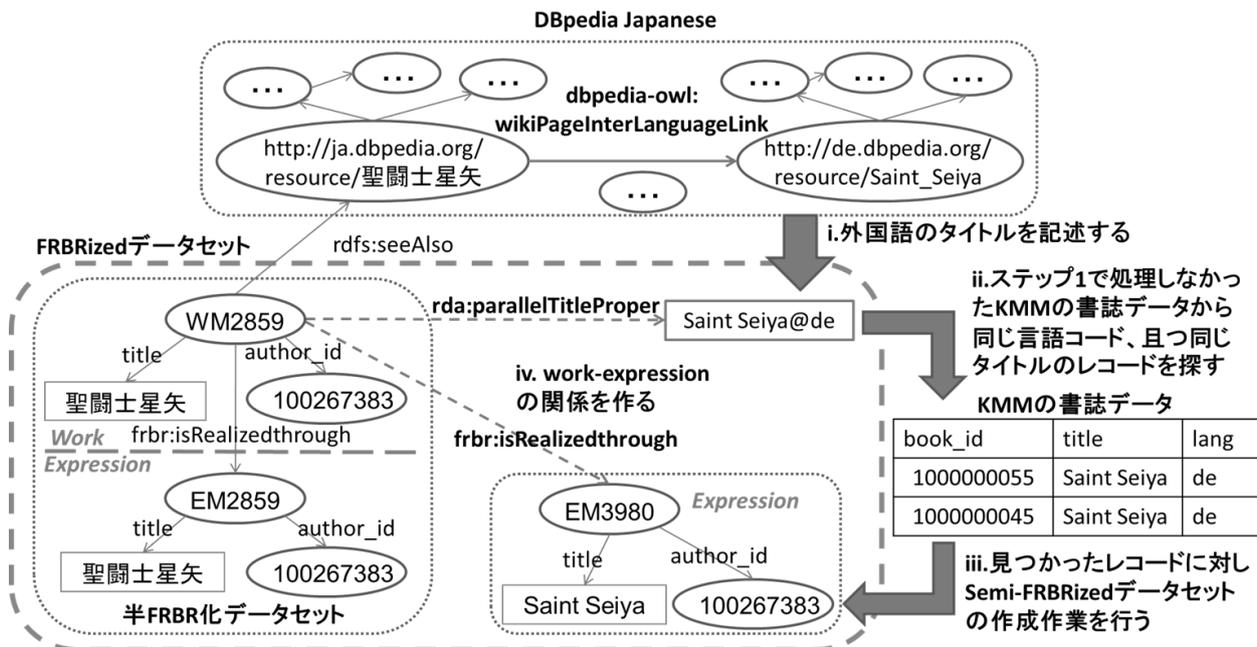


図 6. 1 対 n の work と expression の関連の作成

5. 試行の結果とその考察

本試行では、`work` のインスタンスとして 2684 件の `manga-title`、210 件の `story-title` と 1107 件の `manga-compilation` をそれぞれ同定することができた。また、DBpedia を利用し、図 1 で示した各階層間と階層内の関連を機械的に作成した。4.2 章のステップ 1 を行った結果として、`manga-titles` と `story-titles` 間の関

連「is part of」として 210 件を作成したが、manga-title と manga-compilation 間の関連「is carried in」は作成できなかった。そして、ステップ 2 と 3 を通して、125 件の is carried in の関連を作成した。一方で、各階層間の関連作りとして、FRBRized データセットのうち、3252 件 1 対 1 の work-expression（日本語）の関連及び 749 件 1 対 n の work-expression（多言語）の関連を作成した。表 1 は各マンガ実体における work-expression の関連数を示している。

表 1. 各マンガ実体におく work-expression の関連数

Entities	work:expression (Japanese)=1:1	work:expression (multi-language)=1:n
story-title	210	0
manga-title	2126	558
manga-compilation	916	191

なお同定結果作成した実体とその関連については全て人手による検証を行い、一般的な認知と一致する事を確認した。これにより、本研究で提案した LOD リソースを利用し書誌データからマンガの work を同定する手法は有効である。

今回の試行ではタイトル、翻訳タイトル、掲載情報などの情報の一部は、LOD リソースに明示的に記述されていなかった。work の候補と DBpedia Japanese との関連付けに際し、DBpedia Japanese の一般的な URI の設定パターンに従って、KMM の書誌データに記述されたタイトルから該当するリソースの URI を推定することでマッチングに利用した。しかしタイトルや URI の表記の揺れが存在するため、このマッチングは必ずしも可能ではなかった。このように、work 同定作業の精確さと効率性は利用可能な LOD リソースとその品質に大きく依存する。

この課題を解決するためには、マンガ work の同定作業で利用するデータセットのスキーマを作成する必要がある。更に、Web 上の LOD リソースからこのスキーマに適合するデータを機械的に抽出し、データセットを作成するシステムを開発する必要があると考えられる。

6. 関連研究

書誌データを機械的に FRBR 化する取組研究として、1 章で述べた Hickey と谷口[3]がある。谷口は著者とタイトルの組合せは有効な同定キーであるという原則に基づいて、JAPAN/MARC の階層構造を利用し、個々の階層から著者標目、責任表示、記述タイトル、タイトル標目を抽出した。そして、それぞれの階層レベルで同定キーを生成することで、集成的（構成的）な work を同定することができることを示した。

また、機械的に FRBR 化データされたデータと LOD リソースを関連付ける取り組みが行われている。両者を機械的に関連付ける有効なマッピングキーとして、work の属性である著者、タイトルまたリソースタイプが挙げられる。Takhirov [6]は機械的に work とそれに対応する LOD リソースの実体を関連付けるためのワークフローを提案した。また実験を通じて、FRBR の work 実体が LOD クラウドの中から識別できることを明らかにした。また、Duchateau [7]は FRBRpedia というツールを開発した。FRBRpedia は DBpedia Lookup Service を利用し、予め FRBR 化されたデータセットの work 実体に対し、DBpedia からタイトル、著者とカテゴリが work 実体と一致するリソースの抽出を行う。抽出したものをランク付けることで、work 実体とそれに対応する DBpedia のリソースと関連付ける。

本研究は work の同定をするために LOD リソースを使用する点から、上記の既存研究と大きく相違する。

7. おわりに

本研究は DBpedia を利用しライブラリーが所蔵するマンガの書誌データから work を同定する手法を提案し

た。work の同定作業においてマンガの典拠データとして利用出来る情報は不足している。そのため、LOD リソースである DBpedia をその補足情報として利用して KMM の書誌データから work の同定を試みた。これより LOD リソースを利用し、書誌レコードから work 実体と実体間の関連を同定することが可能であることを示した。しかし work 同定作業の精確さと効率は使われた LOD リソースの品質に依存する。今後の課題として、LOD リソースから取得したデータをマンガの典拠データとして統合するシステムが必要である。また、このシステムは、マンガ work の同定作業だけに使うものではなく、LOD リソースからマンガ work の属性であるキャラクター名や名セリフや舞台などの主題情報を収集する基盤ともなると言える。

参考文献

- [1] IFLA Study Group on the Functional Requirements for Bibliographic Records.2009.Functional Requirements for Bibliographic Records Final Report.http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- [2] Thomas B. Hickey, Jenny Toves.2009.for OCLC Research FRBR Work-Set Algorithm (Version 2.0).<http://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf>
- [3] 谷口祥一.2009.FRBR OPAC 構築に向けた著作の機械的同定法の検証--JAPAN/MARC 書誌レコードによる実験.Library and information science.61.119-151.<http://lis.mslib.jp/pdf/LIS061119.pdf>
- [4] Ayako Morozumi, Satomi Nomura, Mitsuharu Nagamori, Shigeo Sugimoto.2009.Metadata Framework for manga: A Multi-paradigm Metadata Description Framework for Digital Comics.<http://dcpapers.dublincore.org/pubs/article/view/979/952>
- [5] DBpedia, <http://dbpedia.org/>
- [6] Naimdjon Takhirov, Fabien Duchateau, Trond Aalberg.2011.Linking FRBR Entities to LOD through Semantic Matching.Research and Advanced Technology for Digital Libraries Lecture Notes in Computer Science.6966.284-295.<http://liris.cnrs.fr/~fduchate/papers/duchateau-tpdl11-linking.pdf>
- [7] Fabien Duchateau, Naimdjon Takhirov,Trond Aalberg.2011.FRBRpedia: a Tool for FRBRizing Web products and Linking FRBR Entities to DBpedia. In: Proc. of JCDL. <http://dl.acm.org/citation.cfm?id=1998183>