

平成 28 年度メディア芸術連携促進事業 連携共同事業

LOD 技術を用いたマンガ目録や評価情報の高機能化
(多言語化) に関する調査研究 実施報告書

インフォコム株式会社

平成 29 年 2 月

目次

第1章 事業概要	5
1.1 事業の背景	5
1.2 実施内容	6
1.2.1 データ調査	6
1.2.2 データ収集	6
1.2.3 データモデル検討	7
1.2.4 データのリンク	7
1.2.5 検証環境構築	7
1.2.6 検証	7
1.2.7 検討会	7
1.2.8 インタビュー	7
第2章 事業の目的・趣旨	8
2.1 目的・趣旨	8
第3章 実施体制	9
3.1 プロジェクト推進体制	9
第4章 実施スケジュール	10
4.1 スケジュール	10
第5章 実施内容	11
5.1 データ調査	11
5.1.1 国内情報	11
5.1.2 海外情報	13
5.2 データ収集	18
5.2.1 メディア芸術データベース	19
5.2.2 DBpedia	20

目次

5.2.3	MyAnimeList	22
5.2.4	その他	23
5.3	データモデル	24
5.3.1	多言語マンガ情報を表現する上での課題	24
5.3.2	検討	24
5.3.3	採用したデータモデル	26
5.4	データのリンク	31
5.4.1	メディア芸術データベースと DBpedia.....	31
5.4.2	国内情報と海外情報	32
5.5	検証ツール	34
5.5.1	ツールの概要	34
5.5.2	OpenLink Virtuoso	34
5.5.3	InfoLib-LOD	35
5.5.4	RDFviewer.....	36
5.6	検証	37
5.6.1	ユーザーの観点による検証	37
5.6.2	目的観点による検証	39
5.6.3	LOD だからできること観点による検証	39
5.6.3	検討委員による検証コメント	40
5.7	検討会	43
5.7.1	第1回	43
5.7.2	第2回	44
5.7.3	第3回	45
5.8	新たな価値創造についてのインタビュー	47
5.8.1	桶田大介先生インタビュー	47

目次

5.8.2 森川嘉一郎先生インタビュー.....	48
5.8.3 遠藤諭氏インタビュー	48
5.8.4 伊藤遊氏インタビュー	49
第6章 成果	51
6.1 LOD 技術を用いたマンガ情報の高機能化	51
6.2 データモデル	51
6.3 各種情報収集手法	51
第7章 総括	52
7.1 総括	52
付録	53
1.クラス設計書	53

第1章 事業概要

1.1 事業の背景

現在の日本では、マンガはアニメ・ゲームと並び代表的な娯楽として、年齢・性別を問わず認められており、同時に海外での人気も高い文化である。経済産業省のクールジャパン政策の中でもマンガ、アニメのようなコンテンツは、海外へ展開する魅力の1つとして挙げられており、今後も積極的に海外へローカライズされていくと考えられる。

海外に広まったマンガは、様々な地域で楽しまれており、それぞれの地域コミュニティで活発な意見交換がなされている。そのため、国内に流通しているマンガ情報と、海外で流通しているマンガ情報では差異が見られる場合が多い。例えば、国内のマンガ情報では、作品の書誌情報や単行本情報などの基本的な情報や、作品紹介、考察など作品に対する情報が豊富に見られる。一方、海外のマンガ情報を見ても「キャラクター」や「出来事」といった作品の中身や、作品が世に与えた影響などといった内容が見受けられ、国内で流通しているマンガ情報とは異なる情報が豊富に見られた。

これらのマンガ情報を利活用してマンガ情報を高度化することで、今まで見えてこなかった新しい情報を見いだせるのではないかと考えた。例えば、国内の創作者に向けた海外での評価を利用したマーケティング手法や、海外のマンガファンに向けた国内のマンガショーケースなど、今までつながっていなかった情報をつなげることで、新しい価値を提供できる。

しかし、実際に各国の地域コミュニティでの意見などを収集しようとしたとき、容易ではないことが分かった。海外での情報を収集する場合、まずは Wikipedia などでの調査が考えられる。日本のマンガの情報が、海外の Wikipedia の記事になっていることは珍しくなく、Wikipedia の言語間リンク¹などを使用することによって容易に収集できる。しかし、記事の内容は各国で異なっており、感想、評価など、個人の意見については全く記載がない。そのため、Wikipedia 以外のサイトに存在する感想、評価などを探すため、Google 検索などで検索を行うが、言語の壁などもあり、対象のサイトを見つけることは容易ではない。また、対象サイトが見つかったとしても、どの項目名に記載されている情報が「作品名」なのかなど、サイトごとに異なっているため情報の収集も容易ではない。

これらの問題がない、海外のマンガ情報を収集できる仕組みは既存サービスでは存在しなかった。そのため、海外の情報を容易に利活用できる仕組みを検討することは有意義であると考えた。

この仕組みを検討する上で、Linked Open Data²(リンクト・オープン・データ、略名: LOD)技術を利用することとした。LOD 技術を利用して、コンピューターで利用しやすい形で公開することによっ

¹ Wikipedia の他言語版へのリンク。他言語版の同じ主題の記事へのリンクとなっており、ページ左の「他言語版」にリンクが並べられている。

² Linked Open Data は、ウェブ上でコンピューター処理に適したデータを公開・共有するための技術の総称である。従来のウェブが HTML 文書間のハイパーリンクによる人間のための情報空間の構築を目的としてきたことに対応して、Linked Open Data では構造化されたデータ同士をリンクさせることでコンピューターが利用可能な「データのウェブ」の構築を目指している。

第1章 事業概要

て、国内と海外のマンガ情報を結び付けた高度なマンガ情報を利活用することが容易になるためである。これにより、高度化されたマンガ情報の利活用が進み、様々な価値が創造されるのではと考えた。

1.2 実施内容

前述の背景から、国内のマンガ情報と海外のマンガ情報を、LOD 技術を用いて、容易に、有機的に結び付けることにより、新しい情報を見いだすことは意義があると考えた。

本事業ではそれを実現するため、まずはインターネット上に存在する国内と海外のマンガ情報から、どのような情報を収集することができるかの調査を行い、その結果を受けて、新しい情報を見いだすために効果的なデータの選定・収集を行った。次に収集したデータの分析を行い、効果的に国内と海外のマンガ情報を結び付けるためのデータモデルを検討して、LOD 技術を用いたデータベースとして構築した。

その後、データベースに格納されたデータを可視化して調査・検証を行うことで、どのような新しい情報が見いだせるかを検討した。

また、外部有識者を招聘（しょうへい）した検討会、外部有識者に対するインタビューなどを実施し、様々な意見を頂く機会を設けた。

実施内容は以下の 8 つから構成される。

- (1) データ調査
- (2) データ収集
- (3) データモデル検討
- (4) データのリンク
- (5) 検証環境構築
- (6) 検証
- (7) 検討会
- (8) インタビュー

1.2.1 データ調査

データ収集に当たり、国内、海外のマンガ情報の調査を行う。インターネット上に存在するマンガ情報の所在を確認し、サイト名、URL、データ内容を表すカテゴリ、データの内容、情報量、更新頻度などを収集していく。この調査結果から、本事業で使用するデータの選定を行う。

1.2.2 データ収集

データ調査の結果に基づいて選定したデータの収集を行う。データ収集の手法については、継続したデータ利活用を想定して、自動化などを用いた、安定してデータ収集可能な手法を検討・実践していく。

第1章 事業概要

1.2.3 データモデル検討

収集したデータについて調査・分析を行い、国内、海外の様々なマンガ情報を扱うためのデータモデルの検討を行う。本事業では LOD の技術を使用する。

1.2.4 データのリンク

収集したデータ間でどのようにリンクを形成するか検討する。例えば、国内情報として複数のサイトから収集したマンガ情報をまとめる際のリンク形成方法、また、国内情報と海外情報をまとめる際のリンク形成方法の検討を行う。1.2.2 データ収集と同じく、データのリンク形成方法についても、継続してデータを利活用できるよう安定してリンク形成ができる手法を検討・実践していく。

1.2.5 検証環境構築

収集したデータを、1.2.3 データモデル検討で検討したデータモデルに格納し、そのデータを可視化できる検証システムを構築する。ここでは LOD 技術を用いたデータベース (RDF ストア³) を採用する。

1.2.6 検証

検証システムを用いたデータの参照を行い、国内、海外のマンガ情報が有機的に結びついたデータベースから、どのような価値創造ができるかの検証を行う。検証の際、「使用者」、「目的」、「LOD」だからできること、などの観点から実施する。

1.2.7 検討会

国内、海外のマンガ情報を高機能化するためにはどのようなことが必要か、また、どのように利活用していくかを検討するため、メタデータモデル、LOD 技術、マンガ情報の有識者を招聘した検討会を開催し、本事業への提言を行う。

1.2.8 インタビュー

本事業で検討したデータモデル、データベースについて、価値創造や他分野での活用の可能性などについて、マンガ、アニメーション、ゲームなどに造詣が深い有識者に御意見を伺う。

³ RDF ストアとは、RDF で記述された情報を蓄積するデータベース。RDF は Resource Description Framework の略で、Web 上にあるリソースを記述するための統一された枠組みで、特にメタデータについて記述することを目的としている。主語・述語・目的語の 3 つの要素でリソースに関する関係情報を表現し、これをトリプルと呼ぶ。

第2章 事業の目的・趣旨

2.1 目的・趣旨

1.1 事業の背景で述べたとおり、日本の代表的娯楽であるマンガは様々な地域で楽しまれておりそれぞれの地域でファン（愛好家）を中心に意見交換が行われてきた。そこで生まれた情報は、それぞれの地域の文化の影響なども受け、多彩なものとなっている。また、日本国内のマンガ情報と海外のマンガ情報を比較したとき、海外のマンガ情報では「キャラクター」や「出来事」といった作品の中身や作品に関する評価といった内容が豊富なケースが多い。

これらの情報を利活用することで従来にはない高度化された「マンガ情報」を整備できると考えた。しかし、利活用には以下のような課題がある。

一つ目は、情報を持つサイトの特定が容易ではないことである。情報を取得するためには、目的とする情報がインターネット上のどこのサイトに存在するかを特定する必要がある。その方法として Google のような Web 検索エンジンでの検索が考えられるが、幾つかの問題がある。例えば Web 検索エンジンで使用するキーワードを決定するため、取得したい情報で使用されている言語に翻訳した作品名、著者名などを事前に調べる必要がある。また、検索結果で出てきたサイトが目的のサイトであるか否か、検索対象の言語で記述されたサイトを読み取って判断する必要がある。これらは検索者が不得意な言語では容易ではないと考えられる。

二つ目は、検索対象のサイトごとに記載方法が異なることである。目的の情報を持つサイトが特定された後、そのサイトに記載された文章から情報を取得するが、どのような標題でどのような情報が記載されているかを特定する必要がある。例えば、「作品の別名」を表す標題が「Alternative Titles」となっているサイトもあれば「Associated Names」となっているサイトもある。このようにそれぞれのサイトでどのような項目名になっているかを把握しないと正しい情報の取得が難しい。さらに、情報取得者の不得意な言語では、どのような項目名が「作品の別名」を表すか、事前の調査が必要になるであろう。

本事業の目的は、国内のマンガ情報と海外のマンガ情報を有機的に結び付け、上記のような課題がない「高度なマンガ情報」を整備し、新たな価値創造の可能性を検証することである。

さらに、LOD 技術を活用して分散環境におけるデータのつながりを実現することで、従来のデータベースでは成し得なかった新しい価値創造の可能性も検証する。

今回の調査は、マンガ領域を対象に行っていくものであるが、この事業の成果を元に、マンガ領域にとどまらないメディア芸術全般を見通したデジタルアーカイブの在り方についても検討していく。

第3章 実施体制

3.1 プロジェクト推進体制

本事業の代表提案者はインフォコム株式会社であり、共同提案者として筑波大学 知的コミュニティ基盤研究センター 知の環境基盤研究部門と一般社団法人 メタデータ基盤協議会と連携して提案を行った。また、外部有識者を招聘した検討会を開催して本事業への提言を行う。

データ連携を行う国内サイト運営主体及び海外サイト運営主体については、国内サイト、海外サイトが保有しているデータの調査を行う中で対象とする運営主体を決定し、データ連携の協力を依頼するものとする。

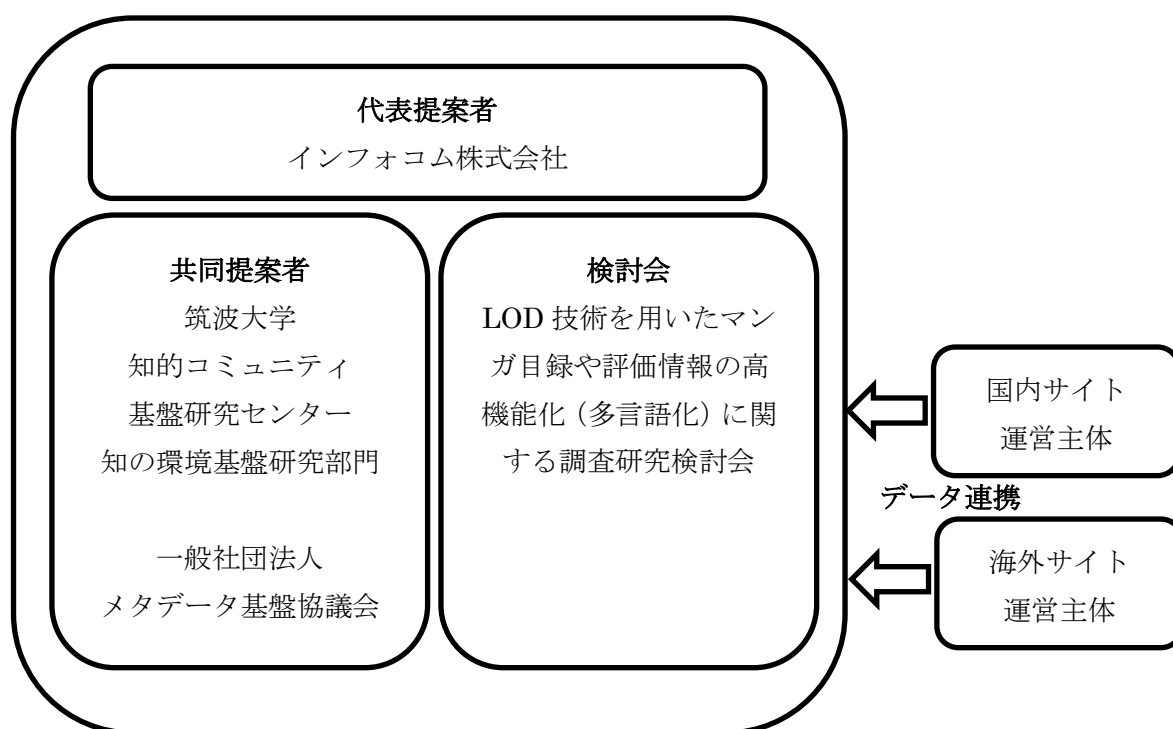


図 3-1 体制図

プロジェクト責任者	: 鳥越 直寿	(インフォコム株式会社)
プロジェクトマネージャー	: 渡部 純	(インフォコム株式会社)
事務局	: 団 宏純	(インフォコム株式会社)
会計担当者	: 武正 一九男	(インフォコム株式会社)
共同提案者	: 永森 光晴	(筑波大学)
共同提案者	: 三原 鉄也	(筑波大学)
検討委員	: 杉本 重雄	(筑波大学)
検討委員	: 武田 英明	(国立情報学研究所)
検討委員	: 池川 佳宏	(株式会社 寿限無)

第4章 実施スケジュール

第4章 実施スケジュール

4.1 スケジュール

実施スケジュールは以下のとおりとする。

表 4-1 スケジュール

	2016年					2017年	
	8	9	10	11	12	1	2
データ調査	■	■	■				
データ収集		■	■	■	■	■	
データモデル検討			■	■	■		
データのリンク				■	■	■	
検証環境構築				■	■	■	
検証						■	
検討会		■			■	■	
インタビュー					■	■	■

第5章 実施内容

5.1 データ調査

本事業で対象とするマンガ情報を選定するため、国内、海外に分けて調査を実施した。

5.1.1 国内情報

・調査概要

本調査は、国内情報として使用するデータ（国内マンガ情報データ）の収集対象を選定するため、日本国内のマンガ情報サイトの状況を把握するために行った。調査内容は、マンガ情報サイトのサイト名、URL、カテゴリ、サイトが所有する情報、更新頻度とした。国内マンガ情報データには、海外のデータをつなげていくハブの役割も期待しているため、情報の網羅度、データ量が重要である。そのため、上記の情報に加え、データ量についても調査を行った。

調査期間は2016年9月1日から10月7日で、方法はGoogle社のGoogle検索を使用したキーワード検索を行い、サイト内容を確認する方法を採用した。

検索キーワードとしては「マンガ」「おすすめ」「評価」「データベース」「ファン」「まとめ」などを用い、検索されたマンガ情報サイトからサイト名、サイトURL、カテゴリ、内容情報、更新頻度、情報量、キーワードなどを収集した。

想定したカテゴリは「ファンサイト（著者・マンガ）」「作品紹介」「評論・考察」「ネタバレ」「Webマンガ」「マンガ投稿」「電子書籍」「データベース」「企業・イベント」とし、各サイトの分類を行った。更新頻度、情報量については正確な情報がサイト上から確認できないため調査者の主観が入ってしまうが、最終更新日時、著者・作品一覧などの情報量を元に判断した。

・調査結果

調査概要にあるように、データのハブとして国内情報の使用を検討しているため、特に情報の網羅度、データ量について調査した。

収集したサイト情報は、全体で108件となった。カテゴリの内訳は以下のとおり。

表 5-1 国内マンガ情報サイトのカテゴリごと件数

カテゴリ	件数
作品紹介	17
評論・考察・ネタバレ	15
電子書籍	13
創作	12
Webマンガ	9
ファンサイト	8
イベント	8

第5章 実施内容

ミュージアムなど	7
データベース	8
その他	11
合計	108

マンガ情報サイトとして多くを占めていたのが、評論・考察・ネタバレ、ファンサイト、そして作品介绍といった特定の範囲を対象としたものであった(40件 38%)。特定の範囲とは、例えば少年マンガ、少女マンガ、バトルマンガといったカテゴリ、特定の雑誌・出版社、特定の著者・作品といったものである。マンガ情報の網羅度を考えたとき、これらのサイトは個人が作成しているものが多く、網羅度は低い。作品介绍サイトにおいて、一部企業が運営しているものも存在するが、自社で出版している作品に限られるため、こちらもマンガ情報の網羅度はそれほど高くない。

企業が運営している電子書籍、イベント、ミュージアムなどのサイトも多く見られた(28件 30%)。これらのサイトは電子書籍の販売、開催したイベントに関する情報提供、ミュージアムなどに所蔵している書誌情報の提供などを行っている。ここでの情報も、当該サイトに関連した情報のみのため、マンガ情報の網羅度は高くないものであった。

オリジナルの Web マンガを発表しているサイト、また、個人が作成したマンガを投稿して公開できるサイトなど、マンガの創作活動に関連したサイトも見られた(21件 23%)。これらのサイトはごく限られたマンガを対象にしており、網羅度はかなり低い。

データのハブとなりえる可能性が高い、データベースとして捉えられるサイトは8件存在した(7%)。データベースのサイトでも一部のサイトは特定の範囲を対象にしたものであったため、マンガ全体を対象としたものは5件であった。情報の網羅度、データ量について、この5件のデータベースサイトはかなり広範囲の作品を網羅しておりデータのハブとして使用できる情報を有していると考えられる。しかし、それぞれのサイトで所有している情報の内容にはそれぞれ特色があった。例えば「メディア芸術データベース」では作品情報、単行本情報、単行本単巻の情報を、「作品データベース」は主に評価、感想を、そして Wikipedia では雑多な情報を所有していた。

・対象サイトの選定

調査結果から、「メディア芸術データベース」、「Wikipedia(DBpedia)」を収集対象とした。それぞれの選定の理由としては以下のとおりである。

「メディア芸術データベース」は情報の網羅度では類を見ない規模であり、書誌情報も充実していたこと。また、情報の信頼性も高く国内のマンガ情報と海外のマンガ情報をつなぐハブとして適していると考えたことなどから選定した。

第5章 実施内容

「Wikipedia」は著者の出生地・得意なジャンル、作品の要約・ジャンルなどメディア芸術データベースにない情報があること、また、「DBpedeia⁴」によりデータ収集が容易であることなどから選定した。

この二つのサイトが所有する情報を組み合わせることで、所蔵情報を含む書誌情報と、作品の要約や著者の出生地などを持つ魅力的なデータとなると考えた。

なお、Wikipedia の信頼性という点では、誰でも編集ができる辞書というところで疑問があるが、Wikipedia 上で編集者同士が議論、編集を繰り返して合意が形成された集合知のため、本事業で使用する上では問題ないと判断した。

5.1.2 海外情報

・調査概要

国内のマンガ情報と結び付けて高機能なマンガ情報を生み出すため、海外のマンガ情報（海外マンガ情報データ）の調査を行った。調査対象国はアメリカ、ドイツ、台湾、タイとし、各国のマンガ情報状況をそれぞれ確認した。

海外マンガ情報データには、国内では見られない外国人の感想、また、キャラクター情報などを期待しているため、感想情報、キャラクター情報が充実しているサイトであるか確認を行った。

・調査結果、対象サイトの選定

A) アメリカの情報について

アメリカの調査対象としては、事前情報として入手できた著名なマンガ・アニメ情報サイトである、「My Anime List」「Manga Updates」「Anime News Network」「AniDB」とした。これらのサイトはマンガ、アニメの情報を豊富にもっており、今回の調査対象としては最適であると考えた。

⁴ DBpedeia は、Wikipedia からの情報抽出を目的とするプロジェクトであり、GPL ライセンスで使用できる。Wikipedia 日本語版を利用した DBpedeia Japanese は、国立情報学研究所によって 2012 年に公開された。

第5章 実施内容

表 5-2 アメリカのマンガ情報サイト

サイト名	URL	特徴	網羅度	評価・感想の充実度
My Anime List	https://myanimelist.net/	<ul style="list-style-type: none"> ・評価、感想が豊富 ・キャラクター情報が豊富 ・日本語での検索が可能 	◎	◎
Manga Updates	https://www.mangauupdates.com/	<ul style="list-style-type: none"> ・カテゴリが豊富 ・評価、感想が豊富 	○	◎
Anime News Network	https://www.animenewsnetwork.com/	<ul style="list-style-type: none"> ・Encyclopedia が充実 ・マンガ制作スタッフ情報 ・アニメの情報が多 ・日本語での検索が可能 	◎	△
AniDB	https://anidb.net/	<ul style="list-style-type: none"> ・アニメに特化 ・日本語での検索が可能 	◎	△

調査の結果、「MyAnimeList」を対象とすることとした。MyAnimeList は過去の手塚治虫作品など 1970 年代の作品情報も所有しており、情報の網羅度では申し分ないと考えた。また、感想、スコア、キャラクター情報の所有量、品質についても、他のサイトと同等、若しくはそれ以上であった。

「MyAnimeList」は日本の株式会社ディー・エヌ・エーの米国子会社である MyAnimeList, LLC が運営しているサイトで、月間 1,000 万人以上が利用している。本事業でのデータ使用については MyAnimeList, LLC より使用許諾を受けた。

B) ドイツ・台湾・タイの情報について

ドイツ・台湾・タイのマンガ情報を持つサイトについては、事前に情報を得ることができなかった。そのため、ドイツ シュツットガルトメディア大学 Magnus Pfeffer 教授、台湾 国立台湾師範大学 Hao-ren Ke 教授、タイ コンケン大学 Kulthida Tuamsuk 教授に情報提供の協力を求めたが、スケジュールの都合により実現できなかった。

上記理由により、Google 検索による情報サイト情報収集や Wikipedia からの情報収集を試みた。

・マンガ情報サイト検索

検索キーワードとしては、「マンガ」「マンガ情報」「マンガイベント」など日本語、ドイツ語、中国語、タイ語などで検索を行った。

第5章 実施内容

表 5-3 海外マンガ情報サイト

国	サイト名	URL	特徴	網羅度	評価・感想の充実度
ドイツ	animexx	http://animexx.onlinewelten.com/	・単行本の情報が豊富 ・レビューは単行本ごと	◎	○
	SPLASH Comics	http://splashcomics.de/php/homepage	・単行本の情報が豊富 ・レビューは単行本ごと	◎	○
	comicguide.de	https://www.comicguide.de/index.php	・単行本の情報が豊富	◎	○
	Carlsen Comics	https://www.carlsen.de/	・出版社サイト	△	×
	Egmont	http://www.egmont-manga.de/	・出版社サイト	△	×
台湾	東立出版社	http://www.tongli.com.tw/	・出版社サイト	△	×
	漫畫博覽會	http://www.ccpa.org.tw/comic/	・イベントサイト	×	×
タイ	NARUTO IN THAILAND	http://board.naruto.in.th/	・NARUTO のファンサイト	×	△
	WEBTOON	http://www.webtoons.com/th/	・NAVER の運営するマンガ掲載サイト	×	△
	NED	http://www.ned-comics.com/	・出版社サイト	△	×

ドイツ・台湾・タイのマンガ情報を持つサイトの検索は難航した。まず、「マンガ」「マンガ情報」「マンガイベント」を翻訳したキーワードで検索しても、目的のサイトが出てくることはまれであった。各国の言語で「マンガ情報」を表す最適な言葉があるかもしれないが、今回の調査では見つけることができなかった。

第5章 実施内容

次に検索できたとしても、台湾、タイでは海賊版などを扱うサイトが出てくることが多かった。日本の出版社が海外企業と連携して海賊版対策に取り組んでいるが、根絶が難しいであろう状況が読み取れた。

また、ドイツのサイトでは作品自体の情報より、単行本を対象とした情報が多く見られた。本事業での海外情報としては、単行本よりも作品に対する感想を取り入れたいと考えていたため、対象サイトとすることを見送った。

どのように海外のマンガ情報サイトを検索するかは今後の課題としたい。

・Wikipedia 情報収集

ドイツ、台湾、タイの Wikipedia からマンガ情報の収集を行った。まず、5 作品を対象に、各国のサイトで収集可能な情報を確認した。まず日本語の Wikipedia を検索し、言語間リンクで各言語の情報にアクセスした。

表 5 - 4 Wikipedia の海外マンガ情報

作品	国	タイトル	要約	単行本情報	ジャンル	出版社 (現地)	キャラクター
AKIRA	ドイツ	○	○	○	×	○	×
	台湾	○	○	×	×	○	×
	タイ	○	○	×	×	×	○
ブラック・ジャック	ドイツ	○	○	○	○	×	△
	台湾	○	○	○	○	○	○
	タイ	○	○	×	○	○	○
キャプテン翼	ドイツ	○	○	○	○	○	○
	台湾	○	○	○	○	○	○
	タイ	○	○	○	○	○	○
ドラゴンボール	ドイツ	○	○	○	○	○	×
	台湾	○	○	○	○	○	○
	タイ	○	○	○	○	○	○
ローゼンメイデン	ドイツ	○	○	○	○	○	△
	台湾	○	○	○	○	○	○
	タイ	○	○	○	○	○	○

調査の結果、Wikipedia から基本的な情報は収集できることが分かった。キャラクター情報も豊富にあり、データ収集対象としては申し分ない。また、ストーリー、分析などの情報もあり、マンガ情報の高度化に寄与できると考える。

第5章 実施内容

しかし、Wikipedia の情報を有効に活用するためには課題がある。それはデータの収集方法である。

前述のとおり、Wikipedia の情報抽出を目的としたプロジェクトとして DBpedia プロジェクトがある。DBpedia を使用することで、LOD 技術を用いて Wikipedia の情報を収集することが容易になるが、取得できる情報は Wikipedia 内の構造化されたデータ (Infobox) がメインとなっている。そこにはストーリー情報やキャラクター情報は含まれていない。そのため、ストーリー情報やキャラクター情報を取得する場合は、Wikipedia 内の本文を解析して収集する必要があるが、Wikipedia 内の本文は、編集者、コミュニティによって記述内容が異なるため、画一的な取得方法を用意することが難しい。例えば、「ドラゴンボール」ではキャラクターの情報は「ドラゴンボールの登場人物」という別の記事となっているが、「AKIRA」のキャラクターは同一記事内に記述されている。また、「ドラゴンボール」の記事内ではマンガ、アニメ、ゲームなどの情報が記述されているため、どの情報がマンガのものかを判断する必要がある。

さらに、該当国に DBpedia がない場合がある。ドイツ、台湾、タイの3か国では、ドイツにのみ DBpedia が存在する。台湾、タイでは基本的な情報を取得する場合でも、Wikipedia 内の本文の解析が必要になってしまう。

Wikipedia の情報を有効活用するためには、上記課題を解決し、できる限り自動的にデータを収集できる環境を用意する必要があると考える。

なお、本事業では、上記調査を行った際、同時に手作業での情報収集を行い、ドイツ、台湾、タイのデータとすることとした。

第5章 実施内容

The image shows a screenshot of the Japanese Wikipedia article for "Dragon Ball". A box labeled "Infobox" points to the right-hand side of the page, which contains the summary and key information about the series. The left side shows the table of contents and a brief overview.

ドラゴンボール

この項目では、鳥山明の漫画について説明しています。その他の用法については「ドラゴンボール (曖昧さ回避)」をご覧ください。

『**ドラゴンボール**』（DRAGON BALL）は、鳥山明（BIRD STUDIO）による日本の漫画作品とその作中に登場するアイテムの名称。略称は『**DB**』。

目次 [非表示]

- 概要
- ストーリー
- 世界観
- 用語
- 背景
- 歴代担当編集者
- 人気
- 単行本・関連書籍
- 関連漫画
- アニメ
- ネット局
- 劇場版
- コンピュータゲーム
- その他のゲーム
- アトラクション
- 脚注
- 外部リンク

概要 [編集]

世界中に散らばった7つの玉を全てを集めると、どんな願いでも1つだけ叶えられるという秘宝・**ドラゴンボール**と、主人公・**孫悟空**（そん ごくう）を中心に展開する「冒険」「バトル」「友情」などを描いた長編漫画である。

『週刊少年ジャンプ』（集英社）にて1984年51号から1995年25号までの約10年半にわたって連載。各話数の通し番号は「其之〇〇」となっており、数字は漢数字で表される。この他に番外編『**TRUNKS THE STORY -たったひとりの戦士-**』（トランクス・ザ・ストーリー -たったひとりのせんし-）も掲載された。

ドラゴンボール DRAGON BALL	
ジャンル	少年漫画 バトル漫画
漫画	
作者	鳥山明
出版社	集英社
その他の出版社	[表示]
掲載誌	週刊少年ジャンプ
発表号	1984年51号 - 1995年25号
巻数	全42巻（JC） 全34巻（完全版）
話数	全519話
アニメ	
・シリーズ作品	
・ドラゴンボール	
・ドラゴンボールZ	
・ドラゴンボールGT	
・ドラゴンボール改	
・ドラゴンボール超	
映画	
・下記参照	
ゲーム	
・下記参照	
<small>テンプレート・ノート</small>	
プロジェクト	漫画・アニメ・ゲーム
ポータル	漫画・アニメ・ゲーム

図 5 - 1 Wikipedia の Infobox

5.2 データ収集

先述のとおり、データ収集対象のサイトとして「メディア芸術データベース」「Wikipedia/DBpedia」、「My Anime List」を選定した。データ収集対象のそれぞれのサイトから、また、ドイツ、台湾、タイのデータとして Wikipedia からデータを収集し、国内マンガ情報と海外マンガ情報を作成する。

まずは「メディア芸術データベース」から情報を収集し、そのデータに「Wikipedia/DBpedia」の情報を付加する。これを国内マンガ情報として活用する。

次に「My Anime List」のデータを、アメリカの海外マンガ情報として収集する。また、ドイツ、台湾、タイの情報は「Wikipedia」のデータを調査中に収集済みである。

なお、データ収集の際、今後の利活用を想定して、安定して継続的にデータを収集できるよう自動化なども念頭に置きつつ手法を検討する。

以下にそれぞれのデータ収集方法についてまとめる。

第5章 実施内容

5.2.1 メディア芸術データベース

メディア芸術データベースは文化庁のメディア芸術デジタルアーカイブ事業の成果の一環として公開されており、マンガ分野では明治初期から2015年12月までのマンガ単行本情報などが登録されている（2016年12月26日現在）。また、マンガ雑誌などの情報も登録されているが、本事業では、まずはマンガ単行本情報のみを対象にデータモデル化することに注力し、マンガ雑誌はスコープ外とした。

情報を収集するに当たり、文化庁にデータ貸与依頼を行い、貸与されたTSVファイルをMS-Accessに取り込み、Ms-Accessの機能（選択クエリー）を利用して必要な情報の取得を行った。選択クエリーでは、対象のテーブルを選択後、データを収集するための条件を入力し、対象とするデータの収集を行った。

その後、選択クエリーを利用して収集したデータをCSVファイルに出力し、更に「Open Refine⁵」を利用してRDFファイルに変換した。

・データ収集手順

1. 貸与されたTSVファイルをMS-Accessにインポートする。
2. 選択クエリーを利用し、取り込んだテーブルから指定した条件のデータを取得する。
3. 取得したデータはCSVファイルで保存する。
4. Open Refineを利用してCSVファイルをRDFファイルへ変換する。

・使用した選択クエリー

使用した選択クエリーの概要は以下のとおりである。

表 5 - 5 MS-Access 選択クエリー概要

クエリー名	概要
作品情報 1	作品の情報を取得するクエリー。マンガ名称、ヨミ、別題・副題・原題、著者典拠idなど。
作品情報 2	作品に紐（ひも）付く単行本全巻のidを取得するクエリー。
単行本全巻情報 1	単行本全巻の情報を取得するクエリー。全巻名、ヨミ、追記、全巻数、出版社名、出版地、言語区分など。
単行本全巻情報 2	単行本全巻に紐付く単行本のidを取得するクエリー。
単行本全巻情報 3	単行本全巻が所属する作品のidを取得するクエリー。
単行本情報 1	単行本の情報を取得するクエリー。単行本名、ヨミ、追記、巻、初版発行年、初版発行月、初版発行日など。
単行本情報 2	単行本が所属する単行本全巻のidを取得するクエリー。

⁵ Open Refine とはデータセットの整形、解析、変換を行うことができるオープンソースソフトウェア。 <http://openrefine.org/>

第5章 実施内容

単行本情報 3	単行本を所有している所蔵館の id を取得するクエリー。
著者情報	典拠の情報を取得するクエリー。名称、典拠 id など。

5.2.2 DBpedia

本事業では、DBpedia Japanese で公開されている SPARQL Endpoint⁶を利用して情報の収集を行い、データの整形などを行った後 RDF に変換している。ここでの RDF 変換でも Open Refine を利用する。

・データ収集手順

- 5.2.1 メディア芸術データベースで作成したデータの著者名、作品名をリスト化する。
例えば、「作品情報 1」クエリーで収集した「マンガ名称」や、「著者情報」クエリーで収集した「名称」などをリスト化する。
- 1 で収集した情報をキーに SPARQL Endpoint にクエリーを発行する。
- JSON データで返戻を受け、取得したデータを CSV ファイルで保存する。
- Open Refine を利用して CSV ファイルを RDF ファイルへ変換する。

・SPARQL Endpoint 情報

以下の SPARQL Endpoint を使用する。

<http://ja.dbpedia.org/sparql>

・使用したクエリー

使用したクエリーの概要は以下のとおりである。

表 5-6 SPARQL クエリー概要

クエリー名	概要
作品要約取得	作品名をキーに「要約」を取得する。 <pre>select distinct ?s ?abstract where { ?s rdfs:label "ドラゴンボール"@ja. optional{ ?s dbpedia-owl:abstract ?abstract. } }</pre> ※上記例ではドラゴンボールの要約が取得できる。

⁶ SPARQL とは、RDF で記述されたデータを検索・操作するための言語である。SPAQL を使って検索・操作するためのインターフェイスが SPARQL Endpoint であり、このインターフェイスを利用することで、外部のアプリケーションから対象の RDF ストアに格納されているデータを検索・操作することができる。

第5章 実施内容

<p>作品ジャンル取得</p>	<p>作品名をキーに「ジャンル」を取得する。</p> <pre>select distinct ?s ?genre where { ?s rdfs:label "ドラゴンボール"@ja. ?s prop-ja:ジャンル ?genre. }</pre> <p>上記例ではドラゴンボールのジャンルが取得できる。</p>
<p>著者情報取得</p>	<p>著者名をキーに「要約」「出生年」「出生地」を取得する。</p> <pre>select distinct * where { ?s rdfs:label "鳥山明"@ja. ?s rdf:type <http://dbpedia.org/ontology/Person>. optional{ ?s dbpedia-owl:abstract ?abstract. } optional{ ?s dbpedia-owl:birthYear ?birthYear. } optional{ ?s dbpedia-owl:birthPlace ?birthPlace. } optional{ ?s prop-ja:生地 ?birthlocalplace. } }</pre> <p>※上記例では鳥山明の要約、出生年、出生地が取得できる。</p>
<p>著者ジャンル取得</p>	<p>著者名をキーに「ジャンル」を取得する。</p> <pre>select distinct * where { { ?s rdfs:label "鳥山明"@ja. ?s rdf:type <http://dbpedia.org/ontology/Person>. optional{ ?s dbpedia-owl:genre ?genre. } } union { ?s rdfs:label "鳥山明"@ja. ?s rdf:type <http://dbpedia.org/ontology/Person>. optional{ ?s prop-ja:ジャンル ?genre. } } }</pre>

第5章 実施内容

	※上記例では鳥山明のジャンルが取得できる。
著者代表作取得	著者名をキーに「代表作」を取得する。 <pre>select distinct * where { ?s rdfs:label "鳥山明"@ja. ?s rdf:type <http://dbpedia.org/ontology/Person>. ?s dbpedia-owl:notableWork ?o. }</pre> ※上記例では鳥山明の代表作が取得できる。

・課題

メディア芸術データベースの情報から DBpedia を検索する際、3点課題があった。

一つ目は作品名の差異である。例えばメディア芸術データベース上では「Rozen Maiden」という作品が、DBpedia 上の検索対象フィールドである「rdfs:label」の値では「ローゼンメイデン」となっているため検索できなかった。しかしながら、他のフィールドの「prop-ja:タイトル」の値には「Rozen Maiden」が存在するため、こちらでの検索も合わせて行う必要があった。また、アルファベット名と片仮名名や中黒・スペースの有無など微妙な表記の違いがある場合がある。この場合は検索時に中黒、スペースを削除してから検索を行うなどの対応が必要となる。

二つ目は著者名の差異である。例えばメディア芸術データベース上では「Peach-Pit」という著者名が、DBpedia 上の rdfs:label では「PEACH-PIT」となっているため検索できなかった。検索時、正規表現マッチなどを使用することで対応は可能である。また、著者名においても、一つ目の作品名の差異と同様に中黒、スペースの問題があるため同様の対応が必要である。

三つ目は Wikipedia の慣例による作品名の差異である。Wikipedia では同名のタイトルが複数存在すると、「作品名 (漫画)」のようなタイトルとなる場合がある。この対応方法としては、二つ目の著者名の差異への対応と同様に、正規表現マッチなどを使用することで対応可能である。

5.2.3 MyAnimeList

本事業では豊富なレビュー、スコア、キャラクター情報を利用してデータの高機能化を行う。データ収集方法としては当初 Beautiful Soup⁷を使用したスクレイピングを検討していたが、Wget⁸ コマンドで対象 Web ページの html ファイルを取得し、html ファイルの内容を目視して手作業で情報収集を実施した。MyAnimeList では「https://myanimelist.net/manga/42/Dragon_Ball」のよう

⁷ Beautiful Soup とは python で動作する HTML と XML のパーサーで、容易な HTML/XML の要素取得を可能にする。<https://www.crummy.com/software/BeautifulSoup/>

⁸ Wget はウェブサーバからコンテンツを取得するダウンローダである。再帰的なダウンロードなど Web クローリングが可能になる機能も持つ。<https://www.gnu.org/software/wget/>

第5章 実施内容

な URL 構造を採用している。この URL の場合「<https://myanimelist.net/manga/42>」としても、同じページにリンクする。そのため、最後尾の数字を変更することでクロールすることができるため、容易に情報収集が可能であった。

情報収集後、CSV に保存して RDF へ変換を行う。

・データ収集手順

1. Wget で myanimelist の web ページをクロールする。
2. 1 で収集した html ファイルから手作業で対象のデータを収集し、CSV ファイルで保存する。
3. Open Refine を利用して CSV ファイルを RDF ファイルへ変換する。

5.2.4 その他

著者の情報を更に高機能化するため、国立国会図書館典拠データ検索・提供サービス⁹ (Web NDL Authorities) を利用し、VIAF¹⁰への URI も取得した。VIAF の URI 取得には、Web NDL Authorities が提供している SPARQL Endpoint を使用した。

・SPARQL Endpoint 情報

以下の SPARQL Endpoint を使用する。

<http://id.ndl.go.jp/auth/ndla>

・使用したクエリー

使用したクエリーの概要は以下のとおりである。

表 5 - 7 VIAF 取得クエリー

クエリー名	概要
VIAF 取得クエリー	典拠 ID をキーに、VIAF への URI を取得する。 <pre>select distinct ?label ?viaf where { ?s <http://xmlns.com/foaf/0.1/primaryTopic> <http://id.ndl.go.jp/auth/entity/00123429>. ?s <http://www.w3.org/2000/01/rdf-schema#label> ?label. optional { ?s</pre>

⁹ 国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities) とは、国立国会図書館で作成し、維持管理する典拠データを一元的に検索・提供するサービス。<https://id.ndl.go.jp/auth/ndla>

¹⁰ VIAF (Virtual International Authority File) とは、国際的な典拠ファイルである。幾つかの国の国立図書館によるプロジェクトで Online Computer Library Center (OCLC) によって運営されている。<http://viaf.org/>

第5章 実施内容

	<pre><http://www.w3.org/2004/02/skos/core#exactMatch> ?viaf. } }</pre> <p>※上記例では典拠 ID : 00123429 (鳥山明) の VIAF URI が取得できる。</p>
--	---

5.3 データモデル

前述したように、本事業では LOD 技術を活用して分散環境におけるデータを有機的につなげることで、新しい価値を生み出そうとしている。ここでは、多言語のマンガ情報を効果的に結び付けることが可能なデータモデルの検討を行った。

以下にマンガ情報を表す方法についての分析を示し、それに基づいてメタデータモデルを提案する。

5.3.1 多言語マンガ情報を表現する上での課題

マンガの利用者がマンガ情報を調べたいとき、タイトル、著者名などで検索することにより一般的な情報を得ることができる。例えば、単行本の書誌情報、所蔵情報、マンガの感想、そして登場するキャラクター情報などを得ることが可能であろう。しかしそれらはそれぞれ独立した情報であることが多く、利用者自身で整理、統合して処理を行わなければならない。

多言語情報で考えたときも上記のような独立した情報の統合が必要なことが多い。例えば「ドラゴンボール」という作品の台湾、タイでの情報を知りたい場合、利用者はドラゴンボールの中国語、タイ語での読み方を調べて検索を行い、中国語、タイ語での情報を自身で読む必要がある。また、タイ在住の外国人が、日本の図書館に所蔵されているマンガ情報を知りたい場合も、日本語の読み方、どのサイトで図書館所蔵情報が分かるか、などの処理が必要になる。

このように、もともと独立した情報の統合が必要な上、更に多言語の情報が独立して存在している場合は様々な壁により情報の収集に困難を来す場合が多い。

上記のようなケースに対応するためには、国内の独立した情報や海外の独立した情報をそれぞれ容易にリンクできるようにしなければならない。

5.3.2 検討

- ・メディア芸術データベースの構造

データモデルの検討をするに当たり、まずはメディア芸術データベースのテーブル構造を参考にすることにした。メディア芸術データベースでは、マンガ作品テーブルが「ドラゴンボール」という作品の概念を表す。マンガ作品テーブルにリンクしたマンガ単行本全巻情報テーブルが単行本全巻という概念を表す。マンガ単行本全巻テーブルにリンクした単行本テーブルが単行本 1 冊ずつを表す。最後にマンガ単行本テーブルにリンクした所蔵情報テーブルが図書館に所蔵された物理的な単行本を表す (図 5-1)。

前述のとおり、本事業ではメディア芸術データベースのマンガ情報をハブとして国内のマンガ情報

第5章 実施内容

と海外のマンガ情報を結び付けようとしている。そのため、メディア芸術データベースの構造をデータモデルとして採用し、国内のデータを格納することにした。

さらに、海外のデータも同じ構造で格納することにした。こうすることにより、言語が異なっても同じラベルが付けられたフィールドは同じ意味のものであると確定できる。例えば、マンガ作品テーブルにある「作品名」、「出版社」に入力されている文字列は、言語が異なっても「作品名」、「出版社名」であると分かる。

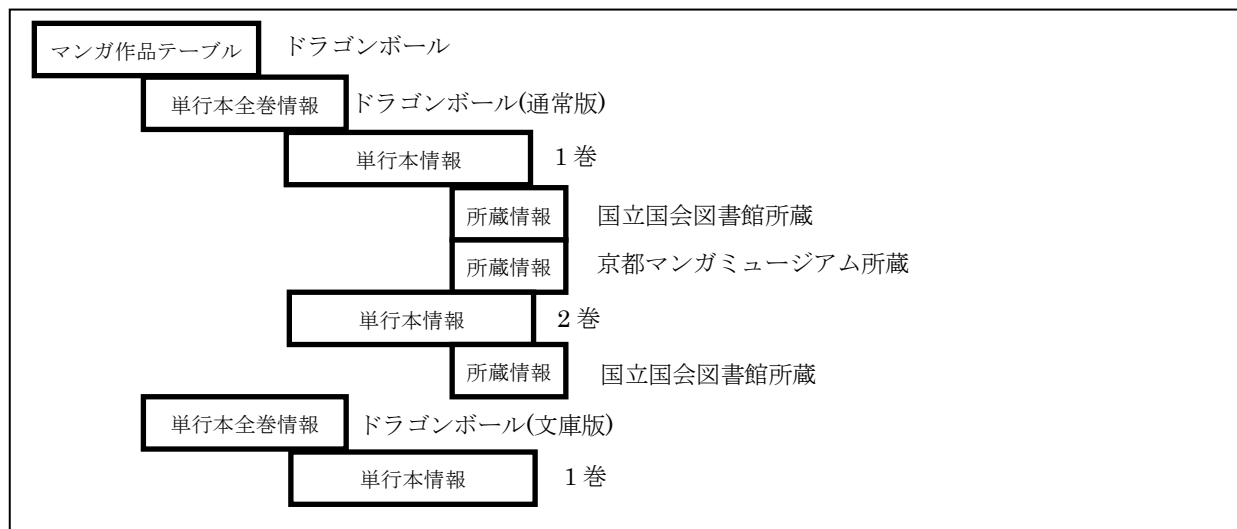


図 5-2 メディア芸術データベース イメージ

・書誌レコードの機能要件：Functional Requirements for Bibliographic Records(FRBR)

次に国内と海外の情報をリンクするに当たって、FRBRを参考にすることにした。FRBRは国際図書館連盟(IFLA)により1997年に勧告された実体関連モデル(ERモデル)を用いた書誌レコードの概念モデルである。書誌レコードに求められる機能について、利用者の観点に立って、明確に定義された用語を用いてモデル化している。

FRBRでは利用者の関心の対象となる実体(entity)が定義され、その第1グループと呼ばれる知的・芸術的な創作である実態には、著作(work)、表現系(expression)、体現系(manifestation)、個別資料(item)として表される。例えば「ドラゴンボール」というタイトルの作品の内容はworkである。workを言語、絵などで表現したものがexpressionである。expressionを紙などで体現したものがmanifestationである。そしてmanifestationが実際に出版されたものがitemとなる(図5-2)。

メディア芸術データベースをFRBRに当てはめて考えてみると、マンガ作品テーブルがwork、単行本全巻情報、単行本情報がexpressionとmanifestationを、所蔵情報がitemを表現していると考えられる。

第5章 実施内容

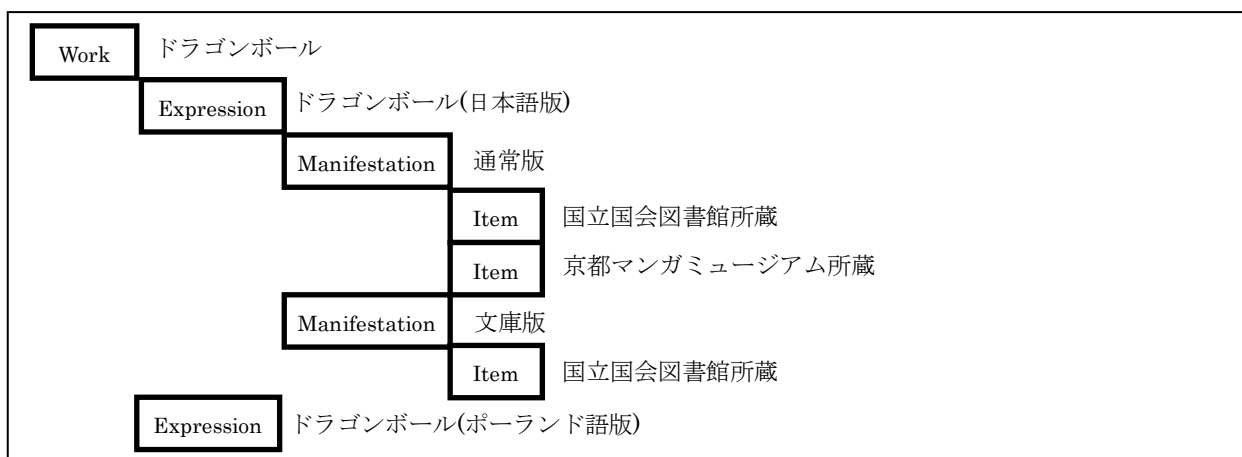


図 5-3 FRBR 第1グループ イメージ

この FRBR の概念を拡張し、各国のマンガ情報を束ねる上位概念を作成し、そこにリンクすることで自由に多言語データを参照できるのではないかと考えた (図 5-3)。この考えは多言語データのリンクにとどまらず、例えば「世界観の同じ作品」、「同じ作品をベースにした作品」など今までになかったつながりを表現するのに適していると考えた。

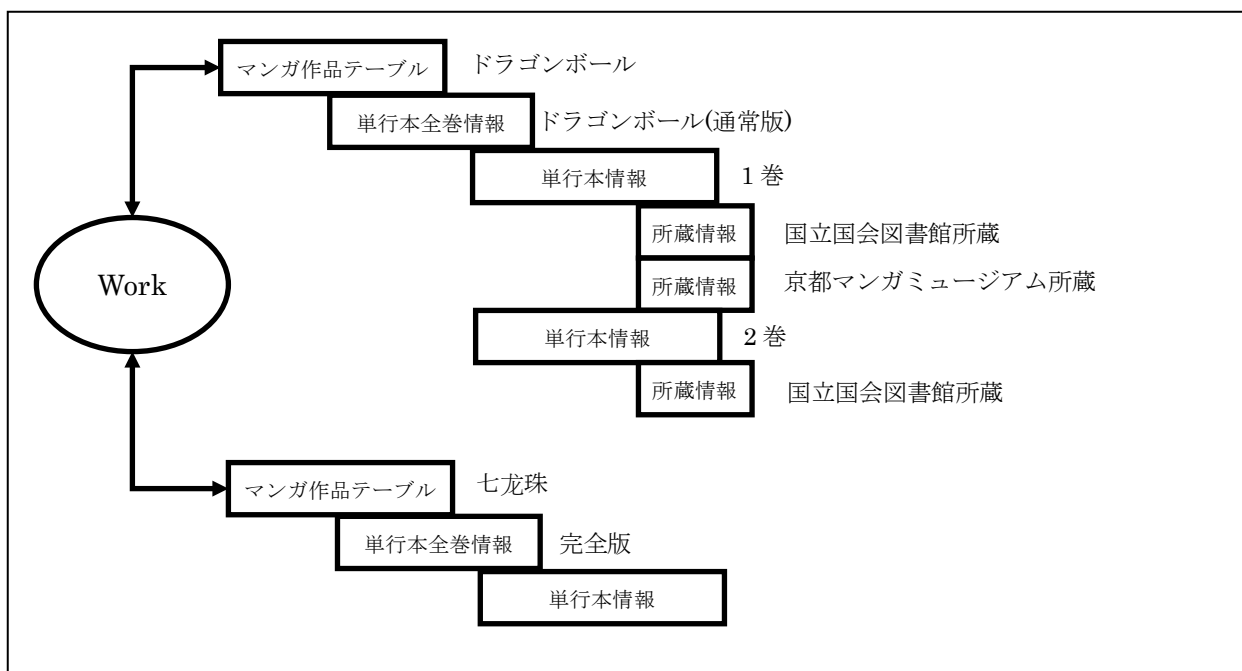


図 5-4 多言語情報のリンク

5.3.3 採用したデータモデル

メディア芸術データベースの構造を元としたデータモデルに FRBR の work をイメージした上位概念を取り入れたデータモデルを採用することとした。

第5章 実施内容

また、メディア芸術データベースの情報を基本としたプロパティ以外にも、マンガ情報の高機能化に寄与すると考えられる情報（マンガの評価情報、キャラクター情報、著者のVIAFリンクなど）の追加も行った。

以下にそのイメージ、設計書の一部を記述する。

1. メタデータ、LOD クラスイメージ

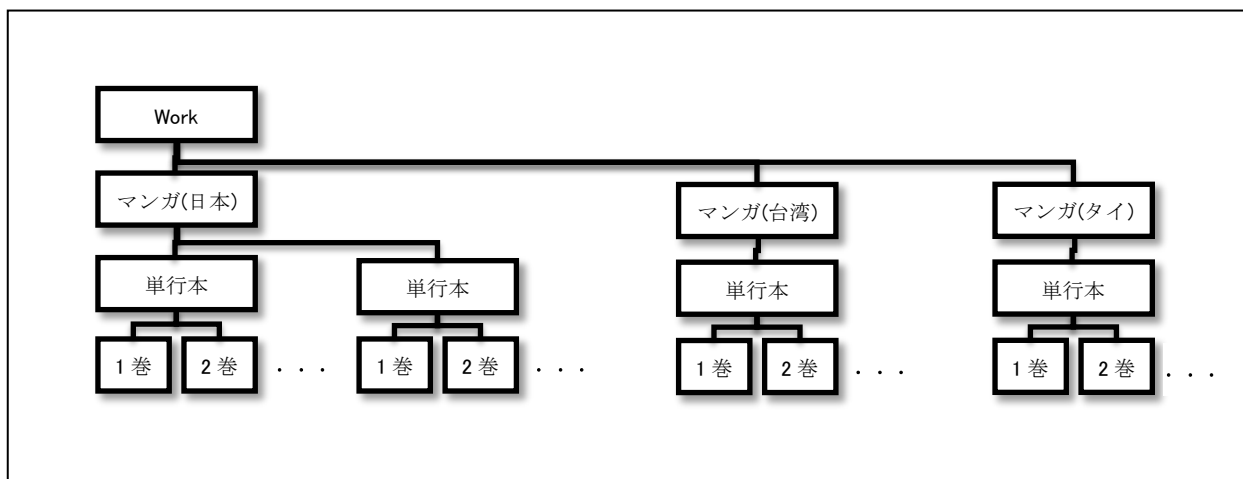


図 5-5 メタデータ イメージ

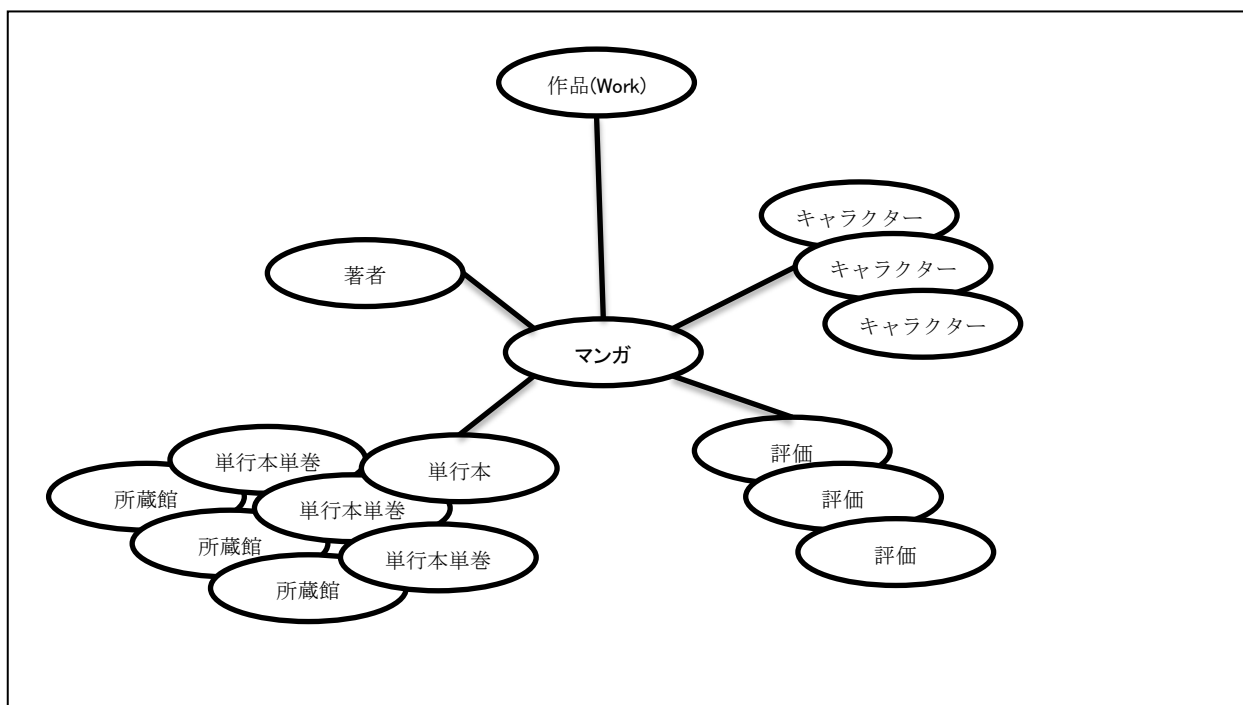


図 5-6 LOD クラス図

第5章 実施内容

2. クラス情報

・概要

表 5-8 クラス概要

クラス名称	概要
作品	work を表すクラス。マンガクラスを結び付ける。
マンガ	マンガ情報を表すクラス。ある地域でのマンガ作品を表す。
単行本	マンガに紐付く単行本を表す。単行本単巻を束ねる概念。
単行本単巻	単行本 1 冊を表す。manifestation を表し、「通常版」、「特典付き版」は別のレコードとなる。
所蔵館	図書館などの館を表す。
著者	著者を表す。著者の情報を持ち、外部へのリンクとなる NDL、VIAF の URI も持つ。
評価	あるマンガに対する評価情報を表す。サイトに掲載されている評価 1 件を 1 レコードとする。
キャラクター	あるマンガに出演するキャラクターを表す。

・作品

表 5-9 作品クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Work
ラベル	作品名称
作品 ID	ID
表現	作品に所属するマンガの URI

・マンガ

表 5-10 マンガクラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Comic
タイトル	マンガの題名
マンガ ID	ID
ヨミ	ヨミ
別題	別名、副題など
著者	著者の URI
ジャンル	ジャンル
説明	説明

第5章 実施内容

評価	評価の URI
キャラクター	キャラクターの URI
単行本	単行本の URI
言語	マンガ情報の言語
地域	マンガ情報の地域
関連するマンガ	マンガの URI
所属する作品	作品の URI

- ・単行本

表 5-11 単行本クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Booktitle
単行本全巻名	マンガの題名
単行本 ID	ID
ヨミ	ヨミ
追記	追記・副題など
全巻数	全巻数
出版社	出版社
出版地	出版地
出版言語	言語
単行本単巻	単行本単巻の URI
所属するマンガ	マンガの URI

- ・単行本単巻

表 5-12 単行本単巻クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Book
単行本名	マンガの題名
単行本単巻 ID	ID
ISBN	ISBN
ヨミ	ヨミ
追記	追記・副題など
巻	巻の番号
出版年	初版出版年
出版月	初版出版月

第5章 実施内容

出版日	初版出版日
所蔵館	所蔵館の URI
所属する単行本	単行本の URI

・所蔵館

表 5-13 所蔵館クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Owner
館名	館名
所蔵館 ID	ID
短縮名	館の短縮名
所有する本	単行本単巻の URI

・著者

表 5-14 著者クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Author
氏名	著者名
著者 ID	ID
著者典拠 ID	NDL の典拠 ID
説明	説明
誕生年	誕生年
出生地	誕生地
ジャンル	ジャンル
代表作品	著者の代表作品
作ったマンガ	マンガの URI
参考情報	NDL、VIAF の URI
別題	別名、別表記（別言語の名称など）

・評価

表 5-15 評価クラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Score
評価 ID	ID
評価者名	評価者名

第5章 実施内容

点数	点数
感想	感想へのリンク
評価サイト	スコアを付けたサイトの URL
点数の最高値	評価サイトでの点数の最高値
評価したマンガ	マンガの URI

- ・キャラクター

表 5-16 キャラクタークラス

名称	内容
タイプ	http://mediam.infocom.co.jp/cs/Character
名称	キャラクター名称
キャラクターID	ID
説明	説明
所属するマンガ	マンガの URI

5.4 データのリンク

各対象サイトから収集したデータを、データモデルに従ってデータベースに格納する際、関連するデータを結び付けるためのリンク情報を形成する。

形成するリンク情報は2種類あり、一つ目はメディア芸術データベースと DBpedia の情報を結び付けるもの。これは、日本語の2つの異なるデータベースを結び付けるリンクである。

二つ目は国内マンガ情報と海外マンガ情報を結び付けるもの。これは日本語と外国語の2つの異なるデータベースを結び付けるリンクである。

これらのリンク情報を作成するため、どのような手法で行うかを以下に示す。

5.4.1 メディア芸術データベースと DBpedia

メディア芸術データベースと DBpedia Japanese は共に日本語を使用するユーザーをターゲットとしたデータベースであるため、名称などが一致する可能性は高い。そのため、単純に同一の名称でリンク情報の形成を行っても問題なく結び付けることができる。しかし、アルファベット名と片仮名名や中黒・スペースの有無など、2つのデータベースで表記が違ふことがあるため、リンク形成時に注意が必要な場合がある。また、Wikipedia の慣例により同一タイトルの記事が存在した場合、「作品名 (漫画) 」のように「 (漫画) 」 「 (アニメ) 」などの文字列がタイトルに付くことがある。この点についても注意が必要である。

以下に、今回行った手法を記述する。

- ・リンク手法

第5章 実施内容

1. メディア芸術データベースの作品名、著者名そのままの値で DBpedia の `rdfs:label` を対象に検索を行う。検索は DBpedia の SPARQL Endpoint に SPARQL クエリーを発行してデータを取得する。以下、データが存在しない場合にキー項目、検索対象項目を変更して検索を続ける。
2. キー項目が中黒・スペースを含む場合、削除する。
3. キー項目がアルファベットの場合、全て大文字、又は小文字に統一する。
4. 上記の検索を、DBpedia の同様の意味を持つ別プロパティを検索対象に行う。例えば「`prop:ja:タイトル`」、「`prop:ja:名前`」、「`foaf:name`」などがそれに当たる。
5. 上記の検索を部分一致検索で行う。

・課題

Wikipedia にはマンガに限らず多彩な記事が存在する。また、マンガの題名も一般的な言葉が使われている例も少なくない。このようなことから、全く異なる記事の情報を取得してしまうことがあった。

例えば大友克洋氏の著作に「ヘンゼルとグレーテル」という作品がある。実はこの作品自体の Wikipedia 記事はないのだが、グリム童話の「ヘンゼルとグレーテル」という記事があり、こちらの情報を取得してきてしまった。この誤りは本文を見ない限り判断がつかない。

本事業では対象作品件数が 458 作品と少数だったため、目視での選択を行ったが、大量データを処理する際など、いかに自動化するかは今後の課題である。

5.4.2 国内情報と海外情報

本事業では、「5.3.3 採用したデータモデル」で前述したとおり、国内マンガ情報と海外マンガ情報を結び付ける情報として「作品 (Work)」を使うこととした。そのデータを作成するため、国内のマンガ情報と海外のマンガ情報を結び付ける方法を検討した。

国内マンガ情報と海外マンガ情報を結び付けるリンクを形成する際、「5.4.1 メディア芸術データベースと DBpedia」で説明した、表記のゆれ、検索対象フィールドの課題もあるが、まずは言語の違いによりリンクが困難になる。海外のサイトで日本のマンガ情報を扱っている場合は日本語の名称を扱っている場合も多いが、国内のサイトで外国語の名称を扱っていることは極めてまれである。特に日本で使われることが少ない言語ではなおさらである。

本事業では、日本語の名称と外国語の名称を対応させる際、Wikidata¹¹の言語間リンクを使用した。言語間リンクは、前述したとおり、Wikipedia のページ左にある「他言語版」の内容で、他言語版の同じ主題の記事へのリンクとなっている。この多言語情報を利用して異なる言語間の名称を同定する。

以下に Wikidata を使用した名称の同定について記述する。

¹¹ Wikidata とは Wikipedia で利用可能な共同編集データベースを提供するプロジェクトである。この成果の一つとして Wikipedia の言語間リンクがある。 <https://www.wikidata.org/>

第 5 章 実施内容

- 名称同定手法
 1. 検索する外国語のキーを用意する。
 2. 1 で収集した情報をキーに Wikidata の SPARQL Endpoint にクエリーを発行する。
 3. JSON データで返戻を受け、取得したデータの `xml:lang` が `ja` の箇所の値を日本語名称として採用する。
- SPARQL Endpoint 情報
以下の SPARQL Endpoint を使用する。
<https://query.wikidata.org/sparql>
- 使用したクエリーの概要は以下のとおりである。

表 5 - 17 Wikidata 言語間リンク取得クエリー

クエリー名	概要
言語間リンク取得クエリー	<pre>select distinct ?label where { ?s rdfs:label "七龍珠"@zh. ?s rdfs:label ?label. }</pre>

- 課題
Wikipedia の記事は各国でそれぞれ編集されている。そのため、国により Wikipedia の記述内容に差異があることから、返ってくる内容が異なる場合がある。例えばドラゴンボールの中国名である「七龍珠」で検索を行った際、日本語名称としては「ドラゴンボール」と「DRAGONBALL EVOLUTION」の 2 レコードが返戻される。これはマンガ・アニメの「ドラゴンボール」とハリウッド映画の「DRAGONBALL EVOLUTION」が中国語ではどちらも「七龍珠」で検索できるためである。現時点では、この状態の解消のためには人の介入が必要で、いかに自動化するかは今後の課題である。

第5章 実施内容

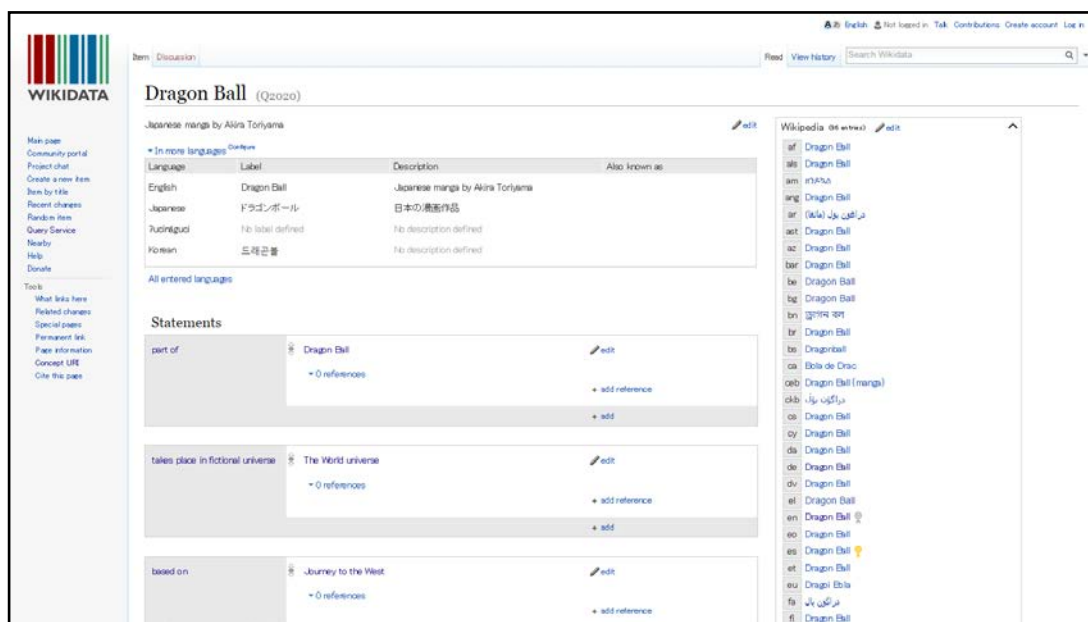


図 5 - 7 Wikidata の Dragon Ball を表すページ

5.5 検証ツール

本事業では、収集したデータ、作成したデータモデルを、LOD 技術を使用して高機能化することにより、新しい価値を創造できることを検証する。そのための各種検証ツールを用意し、データに実際に触れながら検証を行う。検証環境はクラウドサービス上に用意し、検討委員などの検証者ユーザーのみが使用できるよう留意した。

以下に検証環境を記述する。

5.5.1 ツールの概要

本事業では LOD 技術を用いた検証を行うため「OpenLink Virtuoso¹²」、「InfoLib-LOD¹³」、「RDFviewer」を用意した。

5.5.2 OpenLink Virtuoso

以下の InfoLib-LOD、RDFviewer の RDF ストアとして OpenLink Virtuoso を採用した。

¹² OpenLink Virtuoso とは、OpenLink Software 社が提供する RDF を格納して SPARQL Endpoint を経由した共有を可能にするソフトウェア。RDF ストアといわれる。

<https://virtuoso.openlinksw.com/>

¹³ InfoLib-LOD とは、Virtuoso に格納されている RDF を検索し、その結果のデータのつながりを表示するビジュアルライゼーション（可視化）機能を提供するソフトウェア。CSV などからの RDF 変換取り込み機能なども有する。

第 5 章 実施内容

5.5.3 InfoLib-LOD

RDF ストアに格納されているデータを対象に、可視化機能でデータ間のつながりを見るために採用した。

このツールで可視化機能には幾つかの機能があるが、外部データベースへのリンク機能が有用である。この機能は、所有している情報に外部データベースへの URI が含まれていると、そのデータベース内の情報を検索し、一緒に可視化することができる。

例えば「<<http://ja.dbpedia.org/resource/ギャグ漫画>>」のような、マンガのジャンルを表す DBpedia Japanese の URI が含まれていると、そのページに含まれている他の URI も画面に表示する。この例だと「ギャグ漫画」のジャンルが付けられている他の作品などが表示される。この機能により、関連した情報を次々にたどることができ、新しい発見をすることができる。

しかし、次々と外部データベースの情報が表示されてしまうことにより本当に必要なデータが埋没してしまうこともあるため、使用方法には注意が必要である。

The screenshot shows the InfoLib-LOD application interface. The main area displays a network graph with nodes representing categories and titles. The central node is '四月は君の嘘' (April is Your Lie). It is connected to several other nodes, including 'ギャグ漫画' (Gag Manga), '新刊漫画' (New Manga), '青春' (Youth), '学園漫画' (School Manga), '四月は君の嘘Prelude', and '公式ガイドブック四月は君の嘘Prelude'. The nodes are color-coded: blue for categories and red for specific titles.

On the right side, there is a '詳細情報' (Detailed Information) panel with tabs for 'RDF/XML', 'Turtle', and 'JSON'. The selected tab shows the following information:

rdf:type	http://mediam.infocom.co.jp/cs/Work
rdfs:label	四月は君の嘘関連作品
ns3:hasObject	http://mediam.infocom.co.jp/data/comic/129841
ns3:wkid	7

Below the detailed information is a '関連ノード一覧' (Related Nodes List) section with a search input field. The list contains the following items:

- 作品
- 四月は君の嘘
- <http://www.w3.org/2000/01/rdf-schema#Class>
- <http://ja.dbpedia.org/resource/青春>
- 公式ガイドブック四月は君の嘘Prelude
- 新刊漫画
- マンガ
- <http://ja.dbpedia.org/resource/学園漫画>
- 四月は君の嘘
- <http://ja.dbpedia.org/resource/ギャグ漫画>

図 5 - 8 InfoLib-LOD の可視化機能

第5章 実施内容

5.5.4 RDFviewer

前述したように InfoLib-LOD が提供するビジュアライゼーション機能は、データにある外部 URI までリンクした表示を行う。そのことにより新しい発見をすることはできるが、その結果、画面に表示されるデータが多くなりすぎ、目的のデータが他のデータに埋没してしまっていて見えなくなる場合がある。

上記のような状況を回避するため、格納されているデータのみを対象とし、かつ見やすいアイコンで表示するツールを用意した。

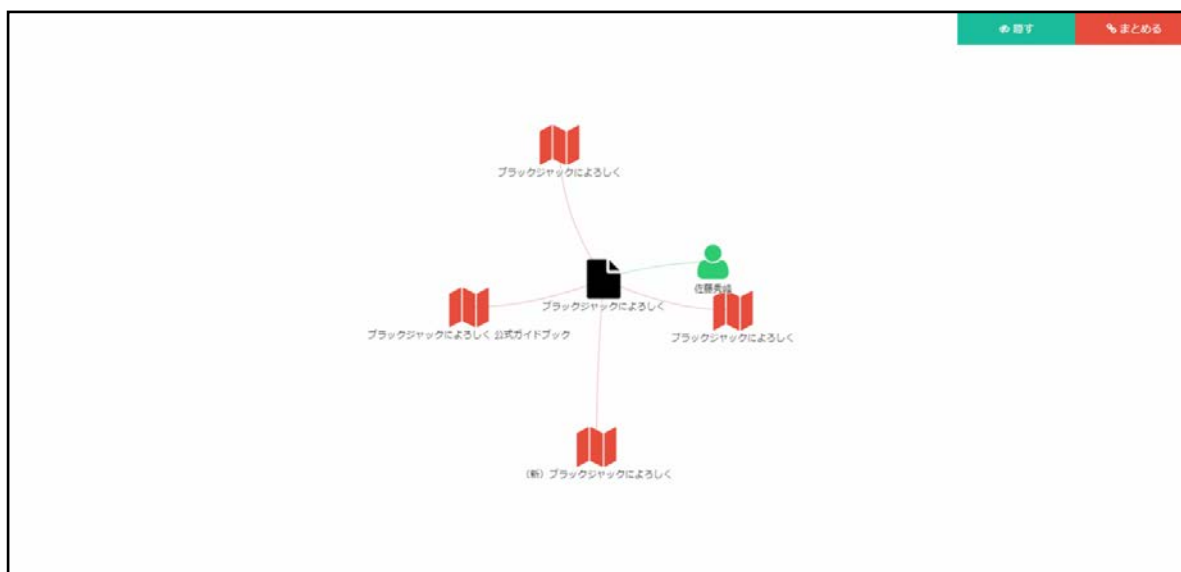


図 5 - 9 RDFviewer 単行本情報

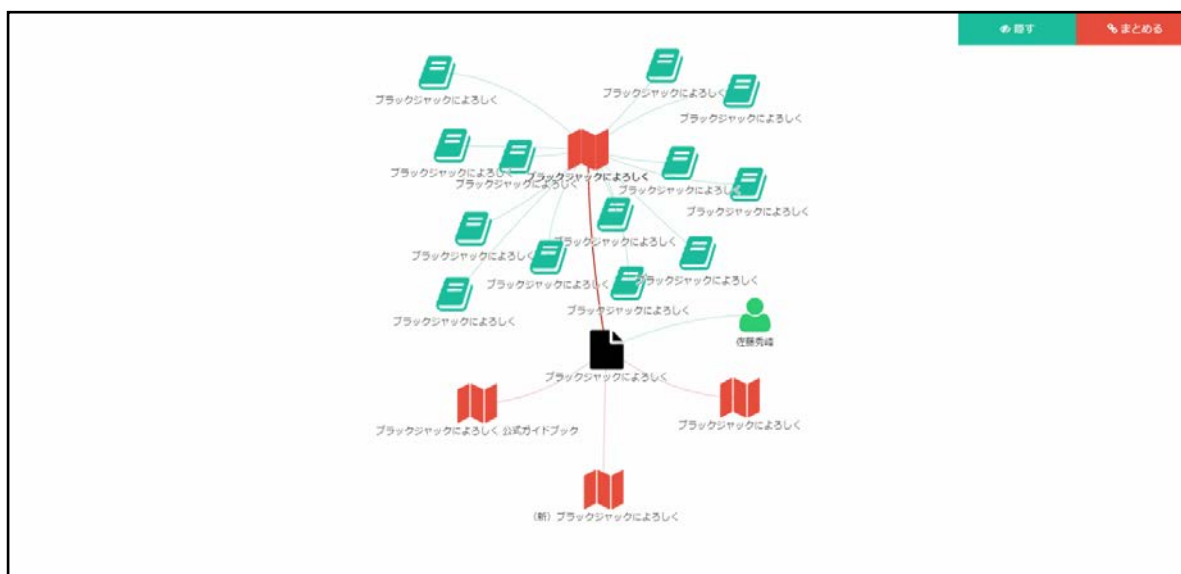


図 5 - 10 RDFviewer 単行本単巻情報

5.6 検証

高機能化された多言語マンガ情報をどのように活用できるか、検証ツールを用いて検証を行うに当たり、幾つかの検証の観点を検討した。

まず一つ目の観点は「ユーザー」である。どのようなユーザーがどのようにデータを使用するか。例えば日本人のマンガファン、又は外国人のマンガ家であればどのような情報が欲しいだろうか。このようなユーザー属性の違いにより求めているものが異なる場合、どのような利活用ができるのか検証する。

二つ目の観点は「目的」である。一般のマンガファンによるマンガ情報収集、同人誌を書いている人が創作活動のため、企業が企業活動に使用するため、など使用目的が異なる場合にどのような利活用ができるのかを検証する。

最後、三つ目の観点は「今まではできないけれども LOD データならできること」である。既存の独立した情報収集ではできない、LOD だからできることにどのようなことがあるのかを検証する。

5.6.1 ユーザーの観点による検証

使用するユーザーの属性の違いによるデータ利活用方法を検証した。

最初は日本人のマンガ家を想定した。多言語のマンガ情報がある場合、自著の外国における評価は興味深く感じるだろう。通常であれば評価、感想が投稿されているサイトを検索し、感想が書かれているページにアクセスすることが難しいが、LOD 技術でリンクされたデータであれば、統一された語彙で登録されているため感想が一目瞭然である。感想に格納されている文字列さえ分かれば、Web 上の翻訳で読み取ることも可能であろう。

次に外国人のマンガファンを想定した。マンガファンであれば、日本のマンガの情報は興味深いだろう。例えば、お気に入りの著者が書いた、日本でしか売っていない作品などの情報が分かる。また、日本に旅行をした際など、作者が生まれた街などへ足を伸ばしたいという希望もあるかもしれない。上記観点での検証を行ったが、希望となるデータの取得は容易であった。

第5章 実施内容

InfoLib-LOD

各レベルの色について

② 「作品(work)」を参照

③ 「マンガ」を参照

④ 「点数」「レビュー」の参照が可能

① 「ドラゴンボール」をキーに「マンガ」が検索される

Dragon ball

ns3:point 10

ns3:scoreId

ns3:belongSite

関連ノード一覧

評価

Dragon ball

http://www.w3.org/2000/01/rdf-schema#Class

マンガ

鳥山明

Dragon ball

ドラゴンボール関連作品

Son Goku

Ten Shin Han

図 5-11 日本語のマンガ情報から海外の評価を参照

InfoLib-LOD

各レベルの色について

③ 「代表作」を参照

④ 「出身地」を参照

② 「著者」を参照

① 「Dragon Ball」をキーに「マンガ」が検索される

Dragon ball

rdfs:label Dragon ball

ns3:language en

ns3:region US

ns3:comicId US00001

http://mediam.infoc

関連ノード一覧

マンガ

鳥山明

Dragon ball

ドラゴンボール関連作品

Son Goku

Ten Shin Han

http://mediam.infocom.co.jp/data/score/1

http://www.w3.org/2000/01/rdf-schema#Class

作品

キャラクター

図 5 - 12 外国のマンガ情報から著者の代表作、出身地を参照

5.6.2 目的観点による検証

使用目的の違いによるデータ利活用方法を検証した。

例えば、マンガ情報の研究を目的とした使用方法を想定した。研究対象のマンガを検索することで、評価、キャラクター情報、関連書籍、所蔵情報などが一目で確認できることが確認できた。

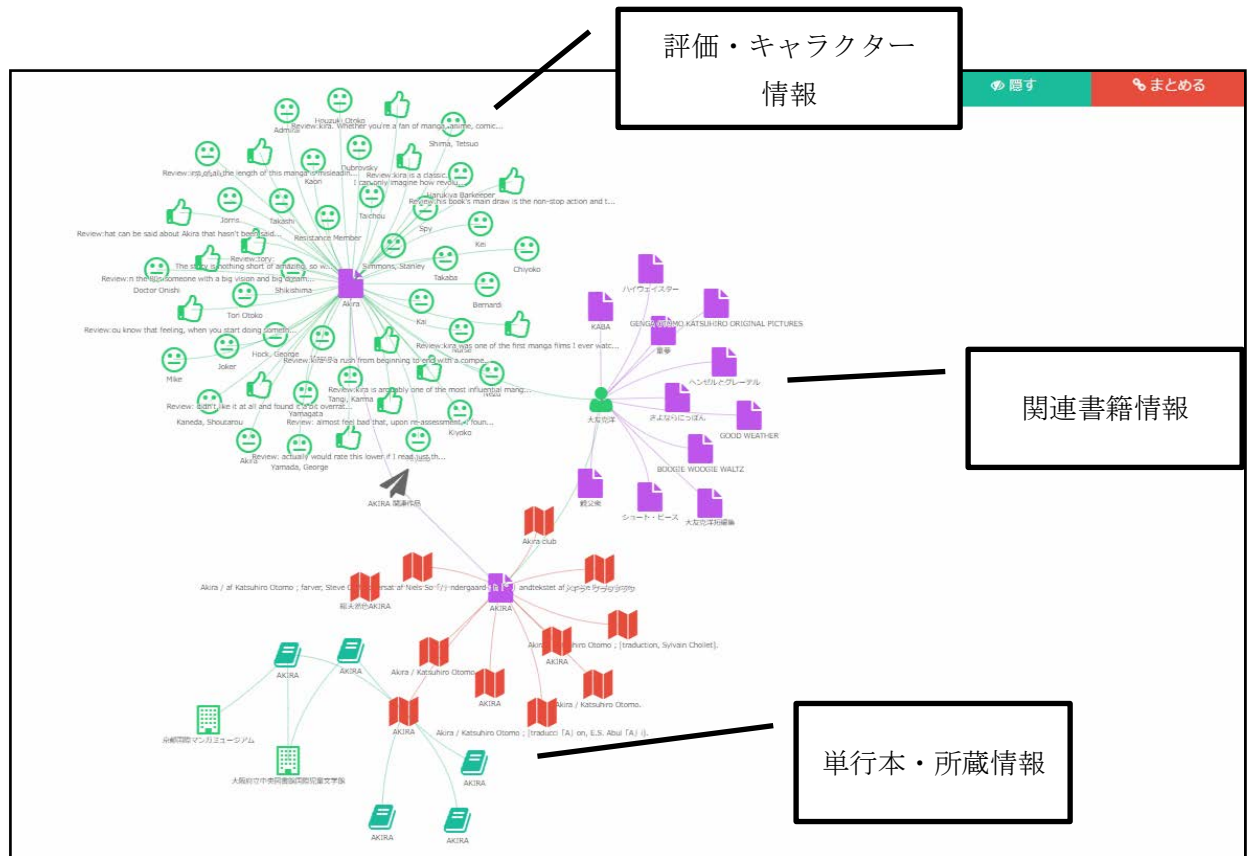


図 5 - 13 マンガ研究のための情報参照

5.6.3 LOD だからできること観点による検証

LOD 技術を用いてマンガ情報を高機能化した結果、本データベースは以下のような特徴をもっている。

一つ目は、語彙の定義が共通化されている。例えば、国内の情報も海外の情報も「作品名称」は「<http://mediam.infocom.co.jp/cs/Comic>」クラスの「`rdfs:label`」プロパティであらわされるし、作品の要約は「`dcterms:abstract`」であらわされる。これは共通のデータモデルに様々なサイトの情報を集めたことにより、各サイトの項目名の違いを吸収することができたためである。

第5章 実施内容

二つ目は、高機能化したマンガ情報を SPARQL で検索することが可能である。これは高機能化したマンガ情報に、外部サイトから SPARQL という統一されたクエリ言語でアクセスすることが可能になったということで、マンガ情報のインフラとして使用できる。

三つ目は、SPARQL で検索することで外部の LOD を大きなデータベースとして使用できる。例えば、外部にある聖地巡礼 LOD に接続することで、作品の聖地情報も扱うことが可能になる。これは外部 LOD によって、使用できる情報が無限に拡大していくことになる。

これらの特徴は、高度化されたマンガ情報を利活用する上で重要なことだと考える。特に、SPARQL Endpoint を利用した外部アプリケーションからの利用は、本データベースをインフラとして利用し、様々な価値を創造することができるのではないだろうか。また外部の LOD データを使用することで無限の情報の広がりがあることは、新しい知見の発見のために有意義であろう。



図 5 - 14 SPARQL Endpoint

5.6.3 検討委員による検証コメント

検討委員にも検証ツールを用いた検証を行っていただき、多言語マンガ情報の価値、LOD 技術を使うことの有用性などについてコメントを頂いた。

①杉本先生

○作成した LOD を用いてデータを可視化することで、作品に対してどういう関連リソースが存在しているのかを俯瞰（ふかん）できる、作品を情報の群として捉えることができる。これは（本取り組みの主眼からは外れるが）作品の分析や比較に応用できるのではと感じた。

第5章 実施内容

○加えて、まだ作品間の関連のデータが不足しているものの、書籍ではなく作品という単位でそれら相互の関係が提示され、**super work** を始め作品を群として俯瞰できることは、従前のキーワード検索やファセット検索にはない作品探索の実現の可能性につながる。

○データをより現実的な作品の探索に利用するためには、データモデルに即した表示機能を充実させる必要があると感じた。具体的には、

- フォーカスしたいクラスの種類によってビュー（特にノード表示）の方法や形式を変える
- クラスによる色分けやフィルタリングなどの機能

といったデータの内容と関連した表示の最適化や明瞭化には工夫の余地がある。

また、そのためには全体的に不十分なプロパティの整備が必要になると考えられる。

○外部リソースとの関連付けも重要な検討課題のひとつであると考えられる。作成した **LOD** は、ファンサイトやオンラインショッピングサイトなどの異なる情報資源を関連づけるためのデータとして利用できるはずである。また、現時点でのデータでは個別のリソースについて、その内容に関する情報が不足しているので、外部リソースを積極的に利用することでそれらを充実させたい。

○**work** のデータについては、**work** の種類や **work** 間の関係をより具体的に検討する必要がある。また、**work** のラベルが同じリソースがたくさんあるのは不適切（サブタイトル(**ns3:addedInfo**)で個別のワークとして識別しているケースが該当する模様）。

○多言語化の観点からは、特定のリソースを選択した際にその他言語の情報を素早く参照できる機能が必要だと感じた。また、言語が異なるだけで対象としては共通のリソースが別々のノードとして提示されるのは紛らわしく、典拠の在り方としても不十分である。**owl:sameAs** プロパティの利用などでインスタンスの集約、典拠化を進めるべきである。

②池川氏

○まず、「メディア芸術データベース」マンガ分野内部のデータ構造がこのような形で可視化されることは大変面白く、これだけでも有効である。

特に、著者経由のつながりは余り「メディア芸術データベース」ではサイト遷移上では重視していないので可視化されてよい。

○さらに、**work** から **MyAnimeList** の情報と関連した「キャラクター」「評価」はメディア芸術データベースにない要素であり、参考情報として有用である。

○次に、サンプル作品を確認した点について。

第5章 実施内容

英語版 wikipedia があっても、work からリンクされていないものが多々ある

藤子不二雄「キテレツ大百科」

https://en.wikipedia.org/wiki/Kiteretsu_Daihyakka

手塚治虫「三つ目がとおる」

https://en.wikipedia.org/wiki/The_Three-Eyed_One

手塚治虫「アドルフに告ぐ」

https://en.wikipedia.org/wiki/Message_to_Adolf

高橋陽一「キャプテン翼」

https://en.wikipedia.org/wiki/Captain_Tsubasa

など。

佐藤秀峰「ブラックジャックによろしく」など、実際に英語版 wikipedia にないものもあるが、「本当は英語版 wikipedia ある」可能性も考えながら確認する必要があった。

キャラクターについて、「きりひと讃歌」の例は恐らく不適切。

手塚治虫「ブラック・ジャック」はスターシステムの一例として「他作品のキャラクターが患者として登場する」場合が多いが、「きりひと讃歌」の reika と「ブラック・ジャック」に登場する reika (wikipedia 英語版による) [https://en.wikipedia.org/wiki/Black_Jack_\(manga\)](https://en.wikipedia.org/wiki/Black_Jack_(manga)) は、キャラクターは造形上も恐らく別と思われる。名前の一致だけでリンクされている事例となっている。

これらの件で、仕組みとしての有効性と、現時点でのリンクの精度についてはまだ乖離 (かいり) があり、課題点が多いと考えられる。

○その他

- ・大友克洋など、海外での評価が高いマンガ家のサンプルがもう少しあるとよい
- ・やはり MyAnimeList はアメリカの情報なので、今後、ヨーロッパ・アジア圏の情報も並行してみられるとよい
- ・単行本の種類が多い作品・マンガ家ほどノードが複雑になるので、可視化の際の調整が難しいところである

第5章 実施内容

5.7 検討会

本事業を進めるに当たり、適切な助言、提言を得るために、有識者を招聘した検討会を開催した。

5.7.1 第1回

開催日時：2016年9月27日 15:30～17:30

場所：インフォコム株式会社

出席者：

【構成員】

- 杉本 重雄 先生（筑波大学）
- 武田 英明 先生（国立情報学研究所）
- 池川 佳宏 氏（株式会社寿限無）
- 永森 光晴 先生（筑波大学）

【文化庁】

- 中臺 正明
- 戸田 康太

【インフォコム株式会社】

- 鳥越 直寿
- 渡部 純
- 団 宏純



図 5-15 第一回検討会

議題：

事業概要の説明を行い、事業の方向性などの検討を行った。また、進めていた国内調査についての中間報告も行い、その内容についての質疑を行った。

トピック：

■Web 上のデータについて

- ・ Web 上のデータを収集した場合、リンク切れなどの問題により維持が問題になる。
- ・ 収集対象が統一されていないので収集が困難では。
- ・ 海外のデータとリンクするための各言語タイトルは **wikidata** も参照材料となる。
- ・ ファンサイトでも信頼できるものはあるが情報量がまばら。
- ・ 主観的な作品像からファンサイトを作ると推測するが、そこから生まれる情報の信頼性は疑問に思う。

■情報基盤

- ・ マンガを 50 年後に残していくコンテンツと考えたとき、作品という概念についても様々な捉え方ができる。概念を決めるのは難しい。コアとなる概念について **URI** が振られるべき。

第5章 実施内容

- ・ LOD で一番欲しいのは作品の安定した Identifier である。作品を同定すること、また、それに対する Identifier を長い時間維持していくためには Wikipedia は大事なリソースなのではないか。
- ・ メディア芸術データベースはタイトル、著者、著者 ID のみで作られている。ID は国会図書館に存在するものはそれを使っている。
- ・ 著作権が一つの課題となるのでは？メディア芸術データベースでは、あらすじやセールスポモーションといった著作権が発生しうるデータは対象外とした。その点を加味すると Wikipedia がたたき台となるのでは。

5.7.2 第2回

開催日時：2016年12月13日 15:00～17:00

場所：インフォコム株式会社

出席者：

【構成員】

- 杉本 重雄 先生（筑波大学）
- 池川 佳宏 氏（株式会社寿限無）
- 永森 光晴 先生（筑波大学）

【文化庁】

戸田 康太

【インフォコム株式会社】

- 鳥越 直寿
- 渡部 純
- 団 宏純

議題：

進捗状況の報告後、以下の議題について検討した。

- ・ 進捗状況報告について
- ・ データモデルについて
- ・ 価値創造について
- ・ アニメ、ゲームなど他分野への適用について

トピック：

■ジャンルについて

- ・ 取扱いが大変難しい。
- ・ 特にこだわらずタグ情報のような形で情報を付与してることが重要。

■FRBR の概念について

第5章 実施内容

- ・メディア芸術データベース、Wikipedia などを見たとき、work からスタートしているものが多い。
- ・item がなく、work からスタートしているため、FRBR は適さないのではないだろうか。
- ・これら work の概念より上位にあるものとして、Super Work というものが考えられる。
- ・work と Super Work の境界は不明確で、Super Work は work でもあると考えられる。

■価値創造について

- ・インバウンド事業は文化政策ともマッチする。
- ・インフラ基盤の必要性を示し、その結果どのようなことが可能になるかを示すのが良いのでは。

■アニメ、ゲームなど他分野へのデータモデル適用について

- ・メディア連携事業の経験から、マンガ、アニメ、ゲームを一つのモデルに取り込んでうまくいくかは疑問。
- ・しかし統一された Super Work は必要なのではないだろうか。

5.7.3 第3回

開催日時：2017年1月30日 15:00～17:00

場所：インフォコム株式会社

出席者：

【構成員】

- 杉本 重雄 先生（筑波大学）
- 武田 英明 先生（国立情報学研究所）
- 池川 佳宏 氏（株式会社寿限無）
- 永森 光晴 先生（筑波大学）

【文化庁】

中村 翔

【インフォコム株式会社】

- 鳥越 直寿
- 渡部 純
- 団 宏純

議題：

最終報告書（案）の説明後、以下の議題について検討した。

- ・最終報告書（案）について

トピック：

■データモデル、検証データについて

- ・どのようにデータを作ったかを残しておくことが重要。

第5章 実施内容

- ・データをマッチさせたときの成功件数、また、工夫をしたのならその後の成功件数の比較なども記載すると良い。
- ・今後のために、自動化できるところ、できないところなどの検討、提案も含めると良い。
- ・FRBRではWikipediaのようなWorkから発生したデータの扱いが難しい場合もある。BIBFRAMEのような形でシンプルにした方が良くもかもしれない。

■Identifierについて

- ・映画にはIMDBのような、かなり信頼のおけるIDがある。しかし他のメディアについてはないのが現状である。
- ・フィルムアーカイブのFIAFという組織があるが、そのような組織があるとIdentifierの割当ても進むのでは。

■Webのみにあるマンガ情報について

- ・今後メディア芸術データベースでも取り込んでいけるよう検討中。
- ・Web上の著作物は消えてしまうといつ発生したかが不明になるので、まずはそこをおさえるようにしたい。
- ・今後、どのWeb著作物を対象とするか、しないかの議論などが必要となってくる。

第5章 実施内容

5.8 新たな価値創造についてのインタビュー

本事業の方向性、実現できる価値などについて、マンガ、アニメーションなどに造詣の深いキーパーソン4名にインタビューを行った。

桶田大介先生（弁護士）：弁護士の他、アニメミライ代表理事、JAniCA 監事、「リトルウィッチアカデミア 魔法仕掛けのパレード」プロデューサーなどアニメをサポートする活動を続けている。

森川嘉一郎先生（明治大学 国際日本学部 准教授）：ヴェネツィア・ビエンナーレ第9回国際建築展で日本館展示のコミッションを務め「おたく：人格＝空間＝都市」と題した展示を企画した。また、「趣都の誕生・萌える都市アキハバラ-」「エヴァンゲリオン・スタイル」などの著作があり、おたく文化、秋葉原の研究活動を行っている。

遠藤諭氏（角川アスキー総研 取締役主席研究員）：サブカル系ライターとして中森明夫氏らと活動し、おたく論、サブカル文化の中心的な存在となった。アスキーに入社後は月刊アスキーの編集長として黄金時代を築いた。著作に「計算機屋かく戦えり」「ソーシャルネイティブの時代 ネットが生み出した新しい日本人」などがある。

伊藤遊氏（京都精華大学 国際マンガ研究センター 研究員）：マンガがどのような人々にどのように読まれているかという「マンガ環境」に関心を持ってマンガ研究を行っている。京都国際マンガミュージアムでは海外マンガに関わる展示やイベントに多く関わってきた。

5.8.1 桶田大介先生インタビュー

2016年12月21日 牛鳴坂法律事務所にて

本事業の取り組みについて、以下のコメントを頂いた。

○アニメーションの典拠になりえる情報について

放送されたアニメーションの情報などの典拠となる情報はない。

しかし、映画になったアニメは「JFDB 日本映画データベース」にある。テレビ局によっては詳細なDBを持っている場合がある。また、EPG（電子番組表）に情報としてはあるはずだが活用できるかどうかは不明。

その他、原口正宏氏らが編纂（へんさん）しているアニメージュの年間パーフェクトデータなどが参考となる。

なお、マンガの典拠に関して言えば、メディア芸術データベースが典拠となりうる。これはNDL OPAC などから情報を取得しているため強度は高い。

○アニメーションに関するアーカイブについて

第5章 実施内容

今までアニメーション業界としてのアーカイブ化は進められていなかったが幾つか動きが出てきている。例えば、日本動画協会のアニメ 100 周年プロジェクトの 1 項目に挙げられている。また、アニメ特撮アーカイブ機構もアーカイブを進めている。

今後、もっと大きな位置づけでのデジタルアーカイブも検討されているのでアーカイブ化は進むのでは。

○価値創造について

聖地情報の提供などに本事業のような仕組みを使うことは、アニメツーリズム協会が設立されるような背景もあり、一定の需要は考えられる。しかし、大きな価値を生むところまでは行くかは未知数。

地域ごとの人気のジャンルや海外の評価・感想から海外に展開する作品を模索する試みは既に行われている。マーケティング主導での創作活動のヒントとはなるかもしれない。

5.8.2 森川嘉一郎先生インタビュー

2017年2月7日 明治大学にて

本事業の取り組みについて、以下のコメントを頂いた。

○海外連携について

海外との連携について協議会識者の関係から取り組むだけでなくネット調査などを活用して網羅的な取り組みで行ってほしい。

○マンガ情報について

世界中のマンガ情報とつながるためには機械的な方策を検討しなければフォローできない。そうした方策を模索しなければ付加価値がない。Google で検索した方が有益となってしまう。Google で世界中の情報を検索し機械翻訳して確認した方がいいということになってしまう。

○検証ツールについて

今回構築した検証ツールが研究者に利用してもらうサービスであるのであれば Google や wikipedia で調査するよりも手間が省ける仕組みである必要があり、また、情報の偏りがあってはいけない。

5.8.3 遠藤諭氏インタビュー

2017年2月7日 角川アスキー総研にて

本事業の取り組みについて、以下のコメントを頂いた。

○有用性及び意義について

海外のマンガ情報と連携する意義はある。以前に Wikipedia のデータを解析したときに「ドラゴンボール」は 47 言語に翻訳されていることが分かった。他のマンガ作品についても幾つもの言

第5章 実施内容

語で翻訳されていることが確認できており、その関心の高さから連携させることに意義がある。ただし、情報を結び付けたサービスだけということでは Wikipedia で実現していることを超えたものがないと有用性が高いとは言い難い。(既存にサービスがあるため)

Wikipedia もひとつの公共サービス。国の基盤で提供することが全てではない。

○データモデル

・分類について

①C コードは編集者と営業で振っている流通のためのコード。「売るため」のコードということでは一定の意味がある。

②本・雑誌の分類に関しては、タイトルや要約・レビュー・作品自体などを元に、単語のベクトル化や深層学習をはじめとする最新の自然言語処理技術を活用すべきと考える。

③コンピューターの書籍の領域ではC コードなどで表現しきれないため、「コンピューター出版販売研究機構」で棚分類コードを作成し、実績を上げてきた。こうした検討も役立つかもしれない。

④出版社ごとに作品作りの手法が異なり、編集者が市場から売れるテーマを模索して作品化していく「製品作り」のようなアプローチもあれば、マンガ家の創造性に依拠した作品作りをすすめるいわば「文芸」的なアプローチもある。前者は分類を付与しやすいが、後者は付与しづらくするのは当然。

とはいえ利用者にとって分類があることは利便性を高めることは間違いない。

・その他

「Linked Open Data」をつかってデータをつないでいきサービスを模索する取組は意義がある。

5.8.4 伊藤遊氏インタビュー

2017年2月3日 京都国際マンガミュージアムにて

本事業の取り組みについて、以下のコメントを頂いた。

○本事業について

マンガへの学術研究においては、マンガ作品そのものを研究対象とすることが多くその関係文献(論文や新聞記事など)を同時に調査することができれば有用。今回の取組でそうした関係文献まで「リンク」が付与されるとより活用できる。

○ジャンル(分類)について

ジャンル(分類)は確立されていないというか、確立するものではない。研究をまとめるに当たり、「ここでは●●を××と定義した上で論考を進める」といった様式となる。

第5章 実施内容

出版社でそれぞれ出版した作品について個別のデータベースを保有しているようなことを聞いたことがある。ただ、そのデータベースにジャンルが明確に付与されているわけではないとのこと。

ジャンルではないが、本取組にはCコード¹⁴と呼ばれる流通のためのコードが使えるかもしれない。Cコードはマンガ単行本に表示されているコード。内容を表すコードも付与されている。

○マンガ業界の関わりについて

データベースとして情報を整理することに対して、出版社及び編集者は関心が薄い。理由は過去ではなく現在を対象に作品作りをするため。ただ、作品の展示会がビジネスにつながるということが分かりつつある中で、全作品のリスト作りといったことに関心が少しずつであるが向かいつつある。

こうした取組では個々の出版社にメリットがある仕組み・仕掛け作りが肝要。検討されている「聖地巡礼」や作品作りの「マーケティング」に有用な取組であると明示化されればより出版社からの協力が得られやすい。

¹⁴Cコードとは図書分類コードとも呼ばれる4桁の数字で、1桁目が販売対象を、2桁目が形態を3・4桁目が内容をそれぞれ表すもの。

第6章 成果

以下に、本事業での成果、及び今後の課題を記述する。

6.1 LOD 技術を用いたマンガ情報の高機能化

国内、海外からマンガ情報を収集し、LOD を用いてマンガ情報の高機能化を行った結果、既存のデータベースでは実現できない新しい情報収集が可能になった。このようなマンガ情報をインフラとして維持していくことで、海外へのマンガ情報発信、国内への海外マンガ情報紹介など、様々な価値が創造されると考える。

しかし、様々なサイトに存在するマンガ情報を、定期的に自動で収集できる方法がなければ、マンガ情報を維持していくことが困難である。さらには、各情報を結び付ける work の情報を誰が、どのように、いつ作成し、どのように維持していくかも大きな課題だろう。

また、本事業では国内のデータをハブとしたため、海外発祥のマンガ情報が対象となっていない。さらに、メディア芸術データベースでは物理的な本からデータが作成されるため、Web のみで公開されているマンガも対象とできていない。これらの情報をどのように収集するかも課題である。

6.2 データモデル

work を1段階上の概念としたデータモデルは、マンガとアニメ、ゲームを結び付ける抽象的な概念として利用できると思う。マンガ、アニメ、ゲームのモデル構造はそれぞれ別だが、work を中心として結び付くことにより同一の概念であるということが示せる。

今後は、更にマンガ情報を高機能化できるプロパティの追加や、アニメ、ゲームそれぞれのデータモデルの検討が必要になってくる。

6.3 各種情報収集手法

本事業では、メディア芸術データベース、DBpedia、My Anime List、また、NDL Authority を用いたデータ収集を行った。それぞれの手法は今後のデータ収集でも利用可能であり、新しい収集方法を探る際の基礎となると考える。

しかし、検索キーワードの微妙な違いや同音異義語により、データの正確性の問題や人の手を介さずに自動で収集できる方法にはなっていない、そもそも、海外のマンガ情報の在りかの検索手法が確立されていない、など課題は多い。また、Blog のような自由に記述されている Web サイトからの情報収集も検討されていない。

前述のとおり、インフラとしてマンガ情報を維持していくためには、自動でデータ収集ができることが大事である。自動的に、大量のデータを、正確に、収集する仕組みを検討する必要がある。

第7章 総括

7.1 総括

本事業では、LOD 技術を用いて多言語データを結び付けることで、データの高機能化が実現できることが示せた。国内のマンガ情報から海外のマンガ情報へのリンクなど、既存のサービスでは見られない新しい結び付きを作ることで、様々な価値創造につながる可能性が見いだせた。

気づいた点としては、データ収集手法など調べてみると、様々な情報が LOD 化されていたため容易に情報が取得できることも多かった。今後も様々なデータベースが LOD 化される可能性は高く、データの取り込みを行わず、全てオープンアクセスでのシステムも検討したい。

今後の改善点として、本事業ではまずは多言語データを用いて LOD の有効性確認を行うことを念頭において進めたため、プロパティの量が少なくなってしまった。Web 上から取得できるデータを更に収集・格納することにより、更なる高機能化が期待できると考える。例えば単行本の ISBN コードなど、単行本の同定に役立つ情報なども Web 上には存在する。これらの有意義な情報も取り入れるようにしたい。

上記のように一定の成果を示せたが、データ収集の自動化、情報の同定方法、効果的な可視化方法など、解決しないといけない課題はまだある。

有識者へのインタビューでも貴重な御意見を頂いた。世界中のマンガ情報を継続的に収集するための方策がなければ絵に描いた餅になってしまうことや、Google、Wikipedia など既存サービスを越えた価値を提供できないと有用性がないことなど、実際にサービスとして提供するためにはまだまだ課題があることが分かった。また、ゲーム、さらにはマンガと近いイメージがあるアニメーションでさえ、マンガが置かれている環境との違いは想像以上に大きく、統一されたモデルを検討する難しさを感じた。

最後に、本事業を通じて、様々な情報を結び付ける work の維持管理がとても重要ではないかという考えを強く持った。誰もが自由に使える uri としての work があると、各自が持っている情報を結び付けることができ、大きな情報となっていく。このようにできた情報は、世界中のあらゆる情報を含み、新たな価値、知見となってくるのではないだろうか。このような環境がインフラとして存在することはとても有意義だと考える。

付録

付録

1. クラス設計書

■クラス定義

名称	クラス URI	上位クラス
作品	http://mediam.infocom.co.jp/cs/Work	rdfs:Class
マンガ	http://mediam.infocom.co.jp/cs/Comic	rdfs:Class
単行本	http://mediam.infocom.co.jp/cs/BookTitle	rdfs:Class
単行本単巻	http://mediam.infocom.co.jp/cs/Book	rdfs:Class
所蔵館	http://mediam.infocom.co.jp/cs/Owner	rdfs:Class
著者	http://mediam.infocom.co.jp/cs/Author	rdfs:Class
評価	http://mediam.infocom.co.jp/cs/Score	rdfs:Class
キャラクター	http://mediam.infocom.co.jp/cs/Character	rdfs:Class

■URI 定義

名称	URI 定義
作品	http://mediam.infocom.co.jp/data/work/(作品 ID)
マンガ	http://mediam.infocom.co.jp/data/comic/(マンガ ID)
単行本	http://mediam.infocom.co.jp/data/bookTitle/(単行本 ID)
単行本単巻	http://mediam.infocom.co.jp/data/book/(単行本単巻 ID)
所蔵館	http://mediam.infocom.co.jp/data/owner/(所蔵館 ID)
著者	http://mediam.infocom.co.jp/data/author/(著者 ID)
評価	http://mediam.infocom.co.jp/data/score/(評価 ID)
キャラクター	http://mediam.infocom.co.jp/data/character/(キャラクターID)

■RDF 語彙(プロパティ)定義

プロパティ URL	ラベル	備考
http://purl.org/dc/terms/abstract	説明	
http://purl.org/dc/terms/alternative	別題	
http://purl.org/dc/terms/language	言語	
http://purl.org/dc/terms/publisher	出版社	
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	分類	
http://www.w3.org/2000/01/rdf-schema#label	ラベル	
http://www.w3.org/2000/01/rdf-schema#seeAlso	参考情報	
http://mediam.infocom.co.jp/element#addedInfo	追記	独自語彙
http://mediam.infocom.co.jp/element#authorId	著者 ID	同上

付録

http://mediam.infocom.co.jp/element#authoritiesId	著者典拠 ID	同上
http://mediam.infocom.co.jp/element#belongBookTitle	所属する単行本	同上
http://mediam.infocom.co.jp/element#belongComic	所属するマンガ	同上
http://mediam.infocom.co.jp/element#belongComic	所属するマンガ	同上
http://mediam.infocom.co.jp/element#belongSite	評価サイト	同上
http://mediam.infocom.co.jp/element#belongWork	所属する作品	同上
http://mediam.infocom.co.jp/element#birthPlace	出生地	同上
http://mediam.infocom.co.jp/element#birthYear	誕生年	同上
http://mediam.infocom.co.jp/element#bookId	単行本単巻 ID	同上
http://mediam.infocom.co.jp/element#bookLanguage	出版言語	同上
http://mediam.infocom.co.jp/element#bookRegion	出版地	同上
http://mediam.infocom.co.jp/element#bookTitleId	単行本 ID	同上
http://mediam.infocom.co.jp/element#characterId	キャラクターID	同上
http://mediam.infocom.co.jp/element#collectedIn	所蔵館	同上
http://mediam.infocom.co.jp/element#comicId	マンガ ID	同上
http://mediam.infocom.co.jp/element#createComic	作ったマンガ	同上
http://mediam.infocom.co.jp/element#createdBy	著者	同上
http://mediam.infocom.co.jp/element#genre	ジャンル	同上
http://mediam.infocom.co.jp/element#hasBook	所有する本	同上
http://mediam.infocom.co.jp/element#hasCharacter	キャラクター	同上
http://mediam.infocom.co.jp/element#hasEvaluation	評価	同上
http://mediam.infocom.co.jp/element#hasObject	表現	同上
http://mediam.infocom.co.jp/element#issuedDay	出版日	同上
http://mediam.infocom.co.jp/element#issuedMonth	出版月	同上
http://mediam.infocom.co.jp/element#issuedYear	出版年	同上
http://mediam.infocom.co.jp/element#maxValue	点数の最高値	同上
http://mediam.infocom.co.jp/element#notableWork	代表作	同上
http://mediam.infocom.co.jp/element#ownerId	所蔵館 ID	同上
http://mediam.infocom.co.jp/element#point	点数	同上
http://mediam.infocom.co.jp/element#publishedAs	単行本	同上
http://mediam.infocom.co.jp/element#reading	ヨミ	同上
http://mediam.infocom.co.jp/element#region	地域	同上
http://mediam.infocom.co.jp/element#relatedComic	関連するマンガ	同上
http://mediam.infocom.co.jp/element#review	感想	同上
http://mediam.infocom.co.jp/element#reviewer	評価者名	同上

付録

http://mediam.infocom.co.jp/element#scoreId	評価 ID	同上
http://mediam.infocom.co.jp/element#separateIntoBook	単行本単巻	同上
http://mediam.infocom.co.jp/element#shortName	短縮名	同上
http://mediam.infocom.co.jp/element#targetComic	評価したマンガ	同上
http://mediam.infocom.co.jp/element#volume	全巻数	同上
http://mediam.infocom.co.jp/element#volumeNumber	巻	同上
http://mediam.infocom.co.jp/element#wkId	作品 ID	同上

■マッピング

作品		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Work
ラベル	rdfs:label	作品名称
作品 ID	mediam:wkId	ID
表現	mediam:hasObject	作品に所属するマンガの URI

マンガ		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Comic
タイトル	rdfs:label	マンガの題名
マンガ ID	mediam:comicId	ID
ヨミ	mediam:reading	ヨミ
別題	dcterms:alternative	別名、副題など
著者	mediam:createdBy	著者の URI
ジャンル	mediam:genre	ジャンル
説明	dcterms:abstract	説明
評価	mediam:hasEvaluation	評価の URI
キャラクター	mediam:hasCharacter	キャラクターの URI
単行本	mediam:publishedAs	単行本の URI
言語	dcterms:language	マンガ情報の言語
地域	mediam:region	マンガ情報の地域
関連するマンガ	mediam:relatedComic	マンガの URI
所属する作品	mediam:belongWork	作品の URI

単行本		
名称	プロパティ	備考

付録

タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Booktitle
単行本全巻名	rdfs:label	マンガの題名
単行本 ID	mediam:bookTitleId	ID
ヨミ	mediam:reading	ヨミ
追記	mediam:addedInfo	追記・副題など
全巻数	mediam:volume	全巻数
出版社	dcterms:publisher	出版社
出版地	mediam:bookRegion	出版地
出版言語	mediam:bookLanguage	言語
単行本単巻	mediam:separateIntoBook	単行本単巻の URI
所属するマンガ	mediam:belongComic	マンガの URI

単行本単巻		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Book
単行本名	rdfs:label	マンガの題名
単行本単巻 ID	mediam:bookId	ID
ISBN	dc:identifier	ISBN
ヨミ	mediam:reading	ヨミ
追記	mediam:addedInfo	追記・副題など
巻	mediam:volumeNumber	巻の番号
出版年	mediam:issuedYear	初版出版年
出版月	mediam:issuedMonth	初版出版月
出版日	mediam:issuedDay	初版出版日
所蔵館	mediam:collectedIn	所蔵館の URI
所属する単行本	mediam:belongBookTitle	単行本の URI

所蔵館		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Owner
館名	rdfs:label	館名
所蔵館 ID	mediam:ownerId	ID
短縮名	mediam:shortName	館の短縮名
所有する本	mediam:hasBook	単行本単巻の URI

著者		
----	--	--

付録

名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Author
氏名	rdfs:label	著者名
著者 ID	mediam:authorId	ID
著者典拠 ID	mediam:authoritiesId	NDL の典拠 ID
説明	dcterms:abstract	説明
誕生年	mediam:birthYear	誕生年
出生地	mediam:birthPlace	誕生地
ジャンル	mediam:genre	ジャンル
代表作品	mediam:notableWork	著者の代表作品
作ったマンガ	mediam:createComic	マンガの URI
参考情報	rdfs:seeAlso	NDL、VIAF の URI
別題	dcterms:alternative	別名、別表記(別言語の名称など)

評価		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Score
評価 ID	mediam:scoreId	ID
評価者名	mediam:reviewer	評価者名
点数	mediam:point	点数
感想	mediam:review	感想へのリンク
評価サイト	mediam:belongSite	スコアを付けたサイトの URL
点数の最高値	mediam:maxValue	評価サイトでの点数の最高値
評価したマンガ	mediam:targetComic	マンガの URI

キャラクター		
名称	プロパティ	備考
タイプ	rdf:type	http://mediam.infocom.co.jp/cs/Character
名称	rdfs:label	キャラクター名称
キャラクターID	mediam:characterId	ID
説明	dcterms:abstract	説明
所属するマンガ	mediam:belongComic	マンガの URI

■使用している prefix

名称	URI
rdf	<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

付録

foaf	<http://xmlns.com/foaf/0.1/>
owl	<http://www.w3.org/2002/07/owl#>
rdfs	<http://www.w3.org/2000/01/rdf-schema#>
dcterms	<http://purl.org/dc/terms/>
mediam	<http://mediam.infocom.co.jp/element#>

本報告書は、文化庁の委託業務として、京都精華大学が実施した平成28年度「メディア芸術連携促進事業 連携共同事業」の成果をとりまとめたものであり、第三者による著作物が含まれています。転載複製等に関する問い合わせは、文化庁にご連絡ください。