

マンガの配列基準としてのWikipediaカテゴリの分析

小平優衣[†]、三原鉄也[†]、永森光晴^{†††}、杉本重雄^{††††}

筑波大学 図書館情報メディア研究科[†]

筑波大学 図書館情報メディア系^{††}

筑波大学 知的コミュニティ基盤研究センター^{†††}

〒305-8550 茨城県つくば市春日 1-2

概要

書籍の一種であるマンガは、書店や図書館が持つ出版社名や著者名等の書誌情報を利用して本棚に配列される。物理的な制約が少なくなるにもかかわらず、デジタル上でのマンガの配列方法も書誌情報を基準にしたものが主である。しかし読者に求められているのは、書誌情報だけでなく主題やキャラクタといったマンガの特徴的な情報を反映した配列である。Web上にはマンガに関する情報が多く存在しており、その情報を利用することでデジタル環境に適した漫画の配列を実現することができると考えられる。本研究ではマンガの配列のために必要な情報を作り出すことを目標に、マンガの情報を多くもつWikipediaを利用するために、Wikipediaのカテゴリ情報を分析した。この分析ではDBpediaのカテゴリ：漫画カテゴリを親とするカテゴリ(全3421カテゴリ)を取り出し、それらをFRBRモデルの実体の概念を利用して分類し、Wikipediaカテゴリのルールや傾向を分析した。本稿では、分析の過程と分析から得たデータについて示す。

キーワード:マンガ, メタデータ, Wikipedia, DBpedia, カテゴリ分析, FRBR

Analysis of the Wikipedia category for the sequence standard of manga

Yui Kodaira[†], Tetsuya Mihara[†], Mitsuharu Nagamori^{††††}, Shigeo Sugimoto^{††††}

Graduate School of Library, Information and Media Studies University of Tsukuba[†]

Faculty of Library, Information and Media Science University of Tsukuba^{††} Research

Center for Knowledge Communities University of Tsukuba^{†††}

1-2, Kasuga, Tsukuba, Ibaraki, 305-8550, Japan

Abstract

Manga or comic book is a popular material for bookstores and libraries. They sort and layout manga books on book shelves primarily using bibliographic information, e.g. publishers and authors. There are many book store and manga information sites which have bookshelf-like interfaces. Those bookshelf-like interfaces do not have physical constraints to layout books unlike those physical bookshelves at bookstores and libraries. Despite this fact, those bookshelf-like interfaces use bibliographic information to layout books in a virtual bookshelf. Bibliographic information is well organized but likely to lose semantic information such as genre of manga, genre of targeted audience, and leading characters and their genre. The goal of this study is to build a structured category vocabulary for manga which can be applied to sort manga books and layout them in a virtual bookshelf. We used DBpedia Japanese, which provides links extracted from a page and category of Wikipedia, in order to build a structured category vocabulary. We collected 3421 categories following subcategory links from Category:Manga of DBpedia Japanese and categorized those categories using FRBR and some related concepts. Then, we analyzed the rules of Wikipedia categories and trends. This paper shows the process to identify the categories of categories and to build a structured vocabulary which can be applied to build a bookshelf-like interface.

Keywords:Manga, Metadata, Wikipedia, DBpedia, Category Analysis, FRBR

1. はじめに

現在マンガは書店や図書館だけでなく、デジタル上でも閲覧や販売の為に扱われる様になった。人が探しやすくするために、書籍はある基準によって並べられる。マンガも物理的に存在する一般書籍であり、本棚に配列される対象である。

デジタル上でのマンガの配列方法は、タイトルや著者名などの書誌情報が主である。しかし新たに読むマンガを探す読者に求められているのは、マンガの内容を反映した配列であり、その理由としてマンガの内容を重視した売り場の展開が書店で見られることが挙げられる。具体的な方法としては、一部のページが見られる様になっている見本を置くことで、試し読みしてから客が購入することができるようにしているというものである。デジタル上でのマンガを配列するための提案として、マンガに関するメタデータの利用を挙げる。本稿ではこの配列を行う為に、マンガに関する情報源としてWikipediaの分析を行う。第2章で実際の本棚やweb上でのマンガの配列方法の比較を行い、マンガの配列に利用したい情報を挙げる。そして第3章ではWikipediaのマンガに関する情報がどういう構造になっているかを分析する。第4章ではWikipediaを利用した参考研究を紹介し、第5章では本稿のまとめと将来的な希望を述べる。

2. デジタル環境におけるマンガの配列とWikipediaの利用

2.1 マンガの配列

Wikipediaの分析に当たって、どのようなルールでマンガの配列が形成されるのかを述べる。まず身近な本棚として、書店と図書館のマンガの書棚を挙げる。

一般的な書店におけるマンガの配列方法は、まず店内で棚の配置があり、棚の中で出版社名による配置が行われている。細かい配列の基準として作者名50音順が利用されている場合が多い。

図書館は日本十進分類法に基づいて書籍を並べるが、マンガはそれ以下の粒度で扱うことができず、マンガという固まりで置かれている。例外的にマンガの所蔵を行っている京都国際マンガミュージアムでは、対象読者傾向別で棚別に分けた後、著者50音順で並べている。

Web上でマンガを扱っているサイトのマンガの配列を考える。Amazon.co.jp[1]などの書籍通販サイトでは新刊順や購入ランキング順が、小学館[]など出版社のWebサイトでは新刊順やタイトルや著者名順、対象読者別が、そして紀伊国屋書店BookWeb[2]などのマンガ配信webサイトでは新着順、ジャンル(テーマ)別が使われている。

読者が読みたいマンガを探す時に、ある一冊だけを探す場合とそうでない場合がある。後者は共通の傾向を持つマンガの群の中で、ある配列基準によって並べられたマンガの中から自分の求める作品を探す場合がある。本棚の配列はこれを可能にしている。

先程述べたような配置を行うためには、マンガの書誌情報だけではなく内容や他のマンガに関する様々な情報を取得する必要がある。その情報源として、マンガの情報が豊富であるという理由からWikipediaを利用することを想定した。Wikipediaのマンガに関するカテゴリは、タイトル別や著者名別もしくは50音順、対象読者別、レーベル、ジャンル(テーマ)別のカテゴリが存在する。またカテゴリとしては存在しないが、マンガ作品の記事から本の発売日を利用することで新刊の表現が可能である。つまりWikipediaの情報を利用することで、既存の配列方法を表現することが可能である。Web上には多くのマンガに関する情報源が存在している。その中で、データとして参照できるWikipediaの情報は、メタデータとしてマンガの配列と組み合わせることが可能であると考えられる。

マンガに関する情報をメタデータとして利用するために、マンガのためのメタデータの形を考える。

2.2 マンガに関するメタデータ

マンガに関するメタデータとは何かを説明する。マンガの情報は様々な形式や規格によって表現される。書誌情報だけでなく、キャラクターやストーリーなどのマンガの要素、場面などの画像に関する情報、展開やコマなどのマンガの構成情報も記述対象である。これらの内容要素をメタデータとして他の情報に関連付けるには、形式や規格を横断して扱うための方法が必要である。ネットワーク環境におけるマンガの発見やアクセス、再利用のためのメタデータフレームワークが Morozumi ら[3]によって提案されている。(図1)

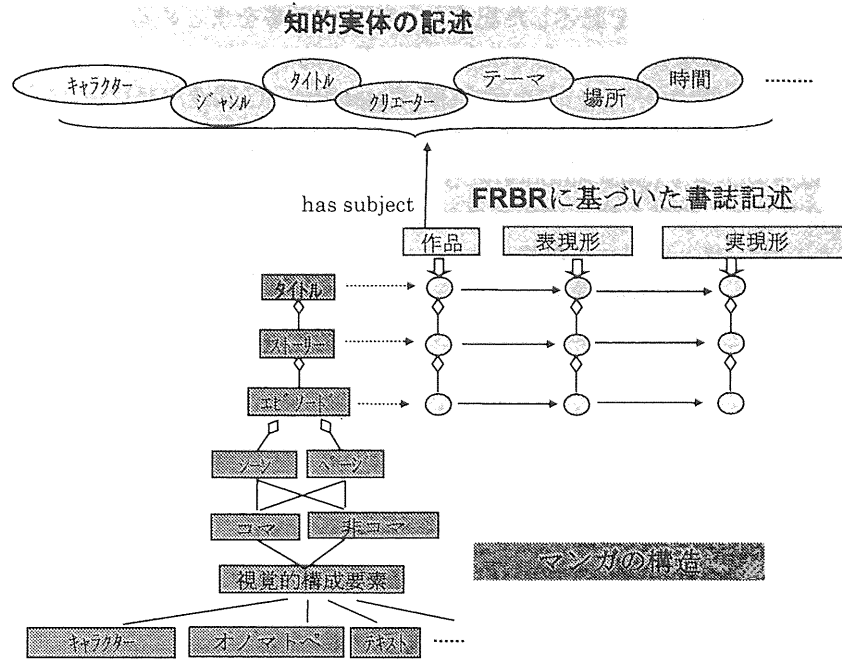


図 1 マンガメタデータフレームワーク [3]

このフレームワークは

- マンガは様々な粒度で記述されることが可能であるべき
- 知的実体としてのマンガと出版された具体物としてのマンガとの差異は明確にされるべき
- 主要人物や動作主体、コンセプトなどのマンガに登場する知的実体はマンガメタデータとしてのオントロジー基盤を利用することで適切に識別、描写されるべき

という要件を踏まえて構成されたマンガメタデータのフレームワークであり、マンガの構造、書誌記述、知的実体の3つの記述に分かれている。「Functional Requirements for Bibliographic Records(以下略称FRBR)」

「TV-Anytime」「知的実体を示す辞書項目」を基礎として拡張構成されたフレームワークであり、その中でもFRBRは書誌をデータの集合体とみなし、実体として捉えるための考え方である。三つのグループを定義しており、第一グループの実体=本そのもの、第二グループの実体=エージェント(作成する、出版する人間)、第三グループの実体=その他本に関係する全ての情報としている。そして第一グループの中で、更に作品(概念そのもの。例:ハリーポッター)、表現系(ハリーポッター日本語版)、実現系(ハリーポッター日本語版第一巻)の三つに分けている。書誌と実体を繋げることで、メタデータとして利用することを可能にしている。

Wikipediaのマンガに関する情報を、このマンガメタデータフレームワークの分け方に沿って分類することで、より整頓された情報として扱うことが可能である。次章でWikipediaのどんな情報がマンガを配列するためのメタデータとして利用することができるかを分析する。

3. Wikipedia内のマンガカテゴリとマンガに関するメタデータの記述

3.1 WikipediaとDBpediaの概要

Wikipediaはインターネット百科事典である。誰もが無料で自由に編集に参加でき、世界の各言語で展開されている。Wikipediaから直接メタデータとして利用可能なデータを取得することはできないため、DBpediaを利用する。DBpediaはWikipediaの情報をLinked Dataとして公開している。DBpediaから取得したWikipediaのマンガに関する情報を用いて分析を行った。

3.2 Wikipediaのマンガに関するカテゴリの分析

現在Wikipediaは対象の詳細情報を記述した記事と、複数の記事をまとめるカテゴリによって構成される。本来のカテゴリは事柄の性質を区分する上でのもっとも基本的な分類のことであるが、Wikipediaのカテゴリは分類だけでなく、関連の深いキーワードを示す(例 Category:○○関連用語)ためにも利用される。またWikipediaのカテゴリはWikipediaの骨格であり、カテゴリの構造はWikipedia全体の構造を示す。そのためにWikipediaカテゴリは子カテゴリ(Subcategory)をもち、下位のカテゴリになるほど扱う粒度が細くなる。マンガに関連する情報として、記事の持つ情報だけでなく、関連の深いキーワードを示すカテゴリを利用することを想定する。しかしユーザーベースの編集により、カテゴリの構造や記事に対するマンガに関する情報の関連づけられ方が不明瞭である。Wikipediaにおけるマンガに関する情報がどういう繋がりや構造をし、どういう要素をもつのかを知るために、カテゴリに注目した分析を行う。

Wikipediaのカテゴリ作成ガイドラインによると、Wikipediaのカテゴリには複数の種類がある。Wikipediaカテゴリは内包する記事が増えすぎたカテゴリを分割するために、カテゴリを利用する(分割として機能するカテゴリ)。しかし全てのカテゴリが分割として機能しているのではなく、分割として機能しないカテゴリも存在する(例 Category:アカデミー賞受賞映画作品に対して「Category:アカデミー賞を受賞しなかった映画作品」は存在しない)。そして関連の深いキーワードを示すカテゴリは、同名の記事を含んでいる場合が多い。カテゴリに内包される記事は、よみがなを基準に50音順でソートされる。

マンガに関する記事とカテゴリは、Category:漫画を最上位のカテゴリとしてみなした。Category:漫画の上位カテゴリはCategory:娯楽、Category:芸術、Category:ジャンル別の書物であり全てのカテゴリに対して最上位ではないが、マンガに関するカテゴリでCategory:漫画を親カテゴリとしてもつカテゴリが多かったためである。Category:漫画に関連する記事の数は約6万件、カテゴリの数は3421カテゴリである。この数からマンガに関する記述の多さがわかる。

Category:漫画から個別の記事に行き着くまでのリンクの一例を図で示す(図2)。楕円はカテゴリを、四角は記事を表現している。実線はカテゴリ同士の繋がりを、点線はページ同士のリンク関係を表現している。

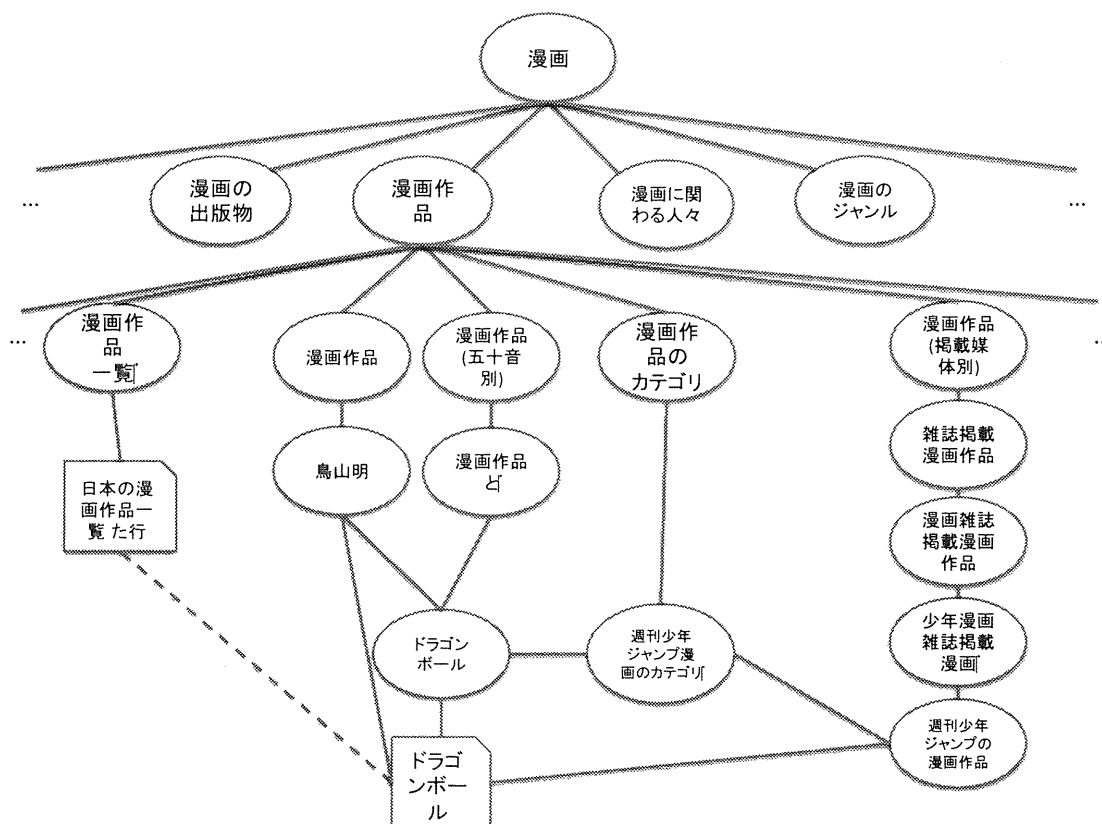


図 2 Category:漫画から個別の記事に行き着くまでのリンク

カテゴリは作品のもつ属性を利用することに有用である。例えば図2のマンガ作品ドラゴンボールは、Category:漫画作品(掲載媒体別)>Category:雑誌掲載漫画>Category:漫画雑誌掲載漫画>Category:少年漫画雑誌掲載漫画>Category:週刊少年ジャンプの漫画作品 に属していることから、作品が漫画雑誌に掲載されていることや、少年漫画であることがわかる。この作品は他のカテゴリにも属しているため、図2以外の属性もカテゴリを辿ることで知ることが可能である。

ここまでWikipediaカテゴリの利用方法について述べてきたが、ここからはWikipediaカテゴリ同士の構造について述べる。Category:漫画以下のカテゴリの繋がりを視覚化した(図3)。

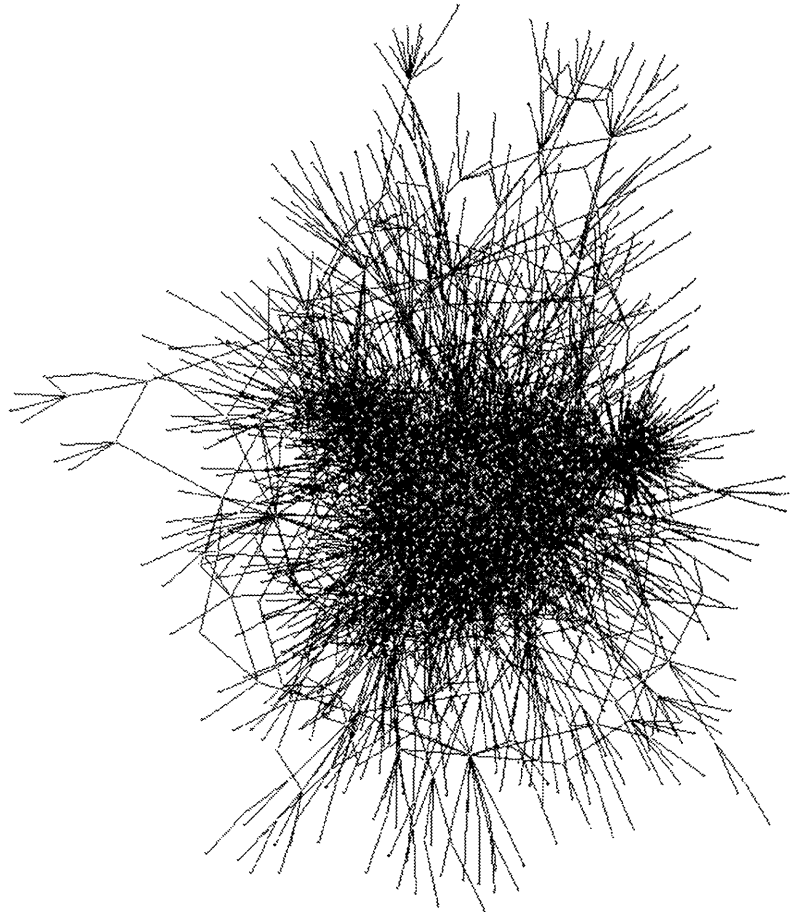


図 3 Category:漫画以下のカテゴリの構造

この図3はカテゴリをノードで表現し、カテゴリ同士の繋がりをエッジで結んで表示したものである。Category:漫画以下に繋がるカテゴリの数は全体で3421カテゴリ(重複含む)存在しており、図3では構造にどのような傾向が現れているのかを確認することは難しい。よってWikipediaカテゴリを分類(カテゴライズ)し、個々に傾向を分析する。カテゴリを分類するための指標としてFRBRを利用した。これによりCategory:漫画以下に繋がるカテゴリを、構造を形成するカテゴリと、マンガ作品とそれに関連するカテゴリに分離することが可能である。これによりCategory:漫画以下の構造を見ることが可能となった。Category:漫画以下に繋がるカテゴリの中からマンガ作品とそれに関連するカテゴリを除いた図を示す(図4)。

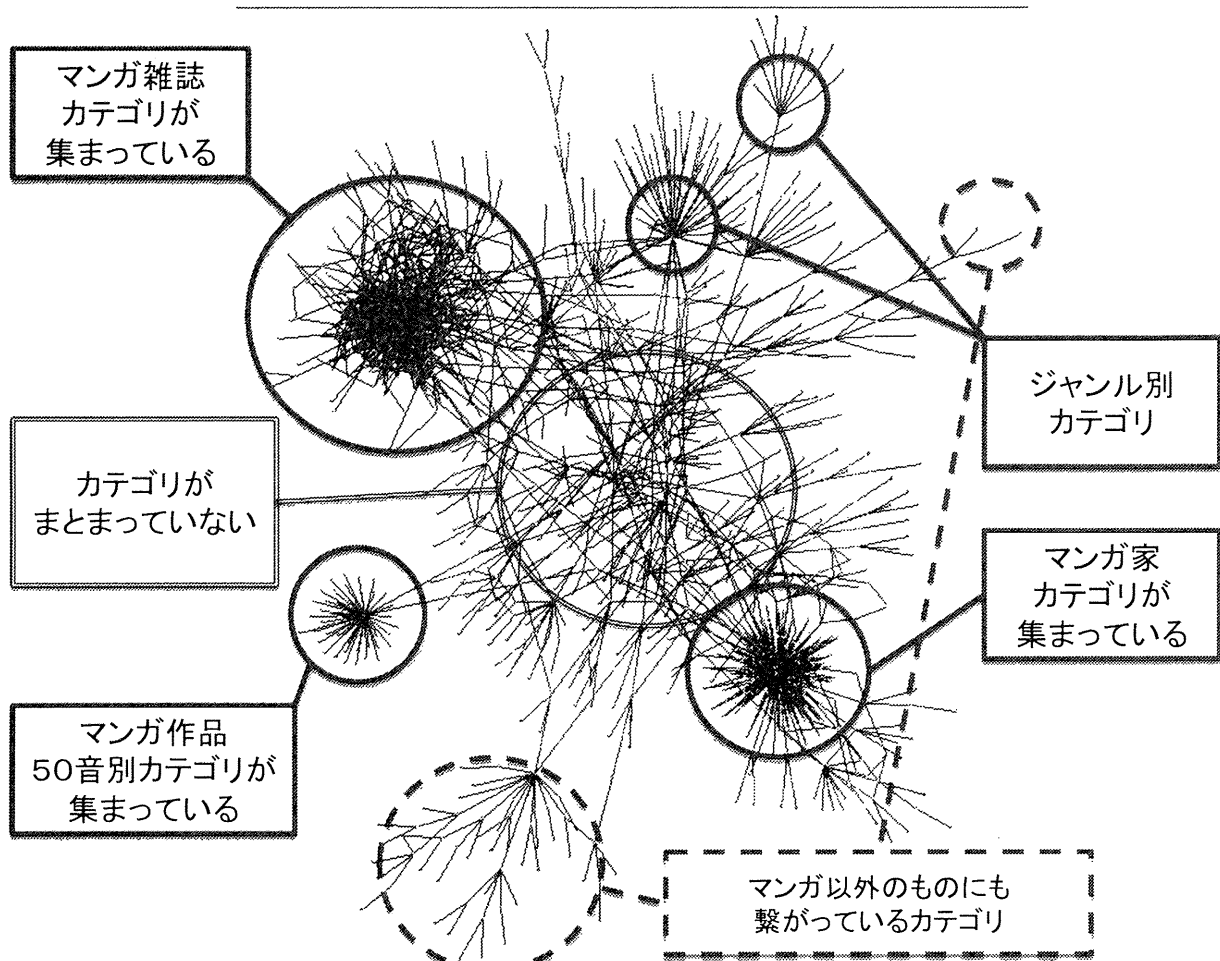


図 4 Category:漫画以下に繋がるカテゴリの構造図からマンガ作品 WORK を取り除いたもの

Category:漫画以下に繋がるカテゴリの構造は、セミ・ラティス構造である。セミ・ラティス構造は、ツリー構造のツリー集合に対して下位と上位の集合の構成要素がそれぞれの集合に含まれている状態で、集合の包含関係が入り組み絡み合っている構造である。例えば、カテゴリは親カテゴリを2つ以上もつ場合がある。

この構造の中でマンガ雑誌のカテゴリ(「Category:週刊少年ジャンプ」など)やマンガ家のカテゴリ(「Category:あだち充」など)が特に多いことがわかる。マンガ家のカテゴリは構造化されていないため、エッジが均等な長さで繋がっている。他には、マンガをテーマや対象読者などのジャンルでわけているカテゴリも複数存在している。またマンガ作品名50音別カテゴリ(「Category:漫画作品_あ」など)は他のマンガに関するカテゴリと関わりを持っておらず、カテゴリ同士の意味を示したものではない。マンガ雑誌カテゴリは多いが、Category:漫画以下のカテゴリは、出版社のカテゴリには直接繋がっていない。それにより「Category:出版社別漫画雑誌」など出版社別にわけるカテゴリが必要となっている。

カテゴリは果たしている役割によって、区別することができる。あるマンガカテゴリのセットを分割するカテゴリだけではなく、例えば「Category:漫画のジャンル」は下に「Category:漫画のジャンル(テーマ別)」

「Category:漫画のジャンル(対象読者別)」 「Category:漫画のジャンル(テーマ別)」を持っている。これらは同じマンガカテゴリのセットを、それぞれテーマと対象読者という別の順で配列しているカテゴリといえる。これら配列のためのカテゴリは、Category:漫画の構造の中に多く見られる。

破線で囲まれた中心から離れているカテゴリは、マンガ以外のものにも繋がっているカテゴリである。例として「Category:ネコを主人公にした物語」はマンガのカテゴリだけではなく、「Category:ネコ科」という生物のカテゴリにも繋がっている。そのため構造の中心から離れてしまう。中心のカテゴリがまとまっていない部分は、多くのマンガに関するカテゴリに繋がっているカテゴリである。例えば「Category:漫画に関わる人々」は「Category:漫画編集者」や「Category:漫画家」や「Category:漫画に関する組織・団体」に繋がっている。

このような多くのカテゴリに繋がっているカテゴリは、多方向からのエッジに引っ張られ、カテゴリ同士でかたまたまにいます。逆にマンガ雑誌のカテゴリは「Category:小学館の漫画雑誌」や「Category: 児童・幼年漫画雑誌」や「Category:月刊漫画雑誌」などの分類のためのカテゴリと「Category:月刊コロコロコミック」などのマンガ雑誌カテゴリが多くのエッジによって繋がっているため、構造を可視化した際に固まりで表示される。

Category:漫画以下に繋がるカテゴリで、構造を形成するカテゴリと分離したマンガ作品とそれに関連するカテゴリについて分析する。本稿では、マンガ作品そのものを示すカテゴリをマンガ作品WORKカテゴリとする。「Category:ドラゴンボール」や「Category:巨人の星」などがこれにあたる。

Category:漫画以下にある全カテゴリ3421カテゴリの内、マンガ作品WORKカテゴリとその下にあるカテゴリは合計で1735カテゴリあり、全体の約50.7%がマンガ作品とその下にあるカテゴリである。その1735カテゴリのうち、マンガ作品WORKカテゴリは975カテゴリ存在する。マンガ作品WORKカテゴリはマンガ作品の中でも複数のカテゴリや記事をもつものに作成され、Wikipediaでは小規模では生成されないポータルページ役割を果たしている。

マンガ作品WORKカテゴリの下のカテゴリとは「Category:(マンガ作品名)の登場人物」や「Category:(マンガ作品名)の音楽」などのことである。複数の種類があるため、下にあるカテゴリを一覧にまとめた。

カテゴリ名	カテゴリ数	マンガ作品WORKカテゴリがこのカテゴリをもつ割合(%)
Category:(マンガ作品名)の登場人物	221	22.7
Category:(マンガ作品名)キャラクタ	25	2.6
Category:(マンガ作品名)のアニメ	55	5.6
Category:(マンガ作品名)のゲーム(作品)	50	5.1
Category:(マンガ作品名)の音楽	108	11.1
Category:(マンガ作品名)アニメソング	44	4.5
Category:(マンガ作品名)の映画(作品)	16	1.6
Category:(マンガ作品名)シリーズ	165	17
Category:(マンガ作品名)一覧	5	0.5

表 1 マンガ作品 **WORK** カテゴリの下にあるカテゴリ

表1から、最もマンガ作品WORKカテゴリの下にあるカテゴリは「Category:(マンガ作品名)の登場人物」だが、マンガ作品WORKカテゴリの数と比較すると、必ず全てのマンガ作品WORKカテゴリが「Category:(マンガ作品名)の登場人物」を下にもっているというわけではない。他のカテゴリも同様である。

4. 関連研究

Wikipediaに関する研究において、カテゴリに着目した研究が行われており、カテゴリやカテゴリ構造の利用方法が多く検討されている。

Wikipediaカテゴリに関する研究として、河内ら[4]はWikipediaカテゴリを利用して検索クエリに関連のあるキーワードを生成する方法を提案している。また田村ら[5]は、既存のWikipediaカテゴリを利用せずWikipediaから抽出した名詞で辞書を作成し、その辞書に基づいてカテゴリを生成しブログを分類している。

カテゴリからオントロジーを構築する研究も多く行われている。Wikipediaのカテゴリを利用し、オントロジーを構築する研究も行われている。柴木ら[6]は、語彙体系クラスの末端階層にWikipediaのカテゴリ構造を対応づけることでオントロジーを構築した。また語彙体系小林ら[7]は、既存のオントロジーとWikipediaのカテゴリを連結することで大規模なオントロジーの構築手法を提案している。

本研究ではかWikipediaカテゴリに着目しているが、Wikipediaの記事二含まれているマンガに関する情報を利用した研究として、村瀬ら[8]はWikipediaの記事本文の概要部分を形態素解析して取り出した単語を利用した作品推薦方法を提案している。

5. 終わりに

本稿では配列のために必要な情報を作り出すことを目標に、マンガの情報を多くもつWikipediaのカテゴリ情報を利用のために分析した。この分析ではDBpediaからCategory:漫画を親とするカテゴリを取り出し、FRBRモデルを利用して分類し、Wikipediaカテゴリのルールや傾向を分析した。その結果を可視化することで、マンガに関するカテゴリの構造がセミ・ラティス構造であり、グラフからマンガ雑誌やマンガ家のカテゴリが多いこと、そして漫画のジャンルカテゴリなどはマンガ作品WORKカテゴリ配列するために作成されているものが多く、そのカテゴリの構造を抽出しマンガ作品カテゴリを並べるだけで配列になるということがわかった。また、マンガ作品WORKカテゴリ以下のカテゴリがマンガに関するカテゴリの半分以上を占めていることや、マンガ作品WORKカテゴリの下にくるカテゴリは決まったものが多いということも判明した。

この分析をもとに、将来的にマンガに関するメタデータを増やし、マンガに関するデータセットの作成や提供を行いたい。またWikipedia以外の語彙（マンガ配信サイトや出版社のマンガ紹介ページなど）と結びつけることで、マンガを配列するために利用したい。

参考文献

- [1] Amazon, <http://www.amazon.co.jp/> (accessed 2013-02-25)
- [2] 紀伊国屋書店 BookWeb, <http://bookweb.kinokuniya.co.jp/index.html> (accessed 2013-02-25)
- [3] Ayako Morozumi, Satomi Nomura, Mitsuharu Nagamori, Shigeo Sugimoto. "Metadata Framework for Manga: A Multi-paradigm Metadata Description Framework for Digital Comics". Proceedings of DC-2009, pp.61-70(2009)
- [4] 河内 沙織, 豊田 哲也, 延原 肇, "Wikipedia カテゴリおよび自己相互情報量に基づく関連検索キーワード生成による知識拡充支援", 情報処理学会全国大会講演論文集, pp.607-609(2011)
- [5] 田村 直之, 伊藤 直之, 西川 侑吾, 中川 修, 新堀 英二, "Wikipediaから作成した辞書によるブログのカテゴリ分類", 情報科学技術フォーラム講演論文集, pp.273-274(2009)
- [6] 柴木 優美, 永田 昌明, 山本 和英, "日本語語彙大系を用いた Wikipedia からの汎用オントロジー構築", 情報処理学会研究報告, pp.1-8(2009)
- [7] 小林 暁雄, 増山 繁, 関根 聡, "Wikipediaと汎用シソーラスを用いた汎用オントロジー構築手法", 電子情報通信学会論文誌, pp. 2597-2609(2010)
- [8] 村瀬 尊好, 終 和佑, 安藤 友晴, "マンガの概要に基づく作品推薦システム", 情報科学技術フォーラム, pp.319-325(2012)