

# コミックデータ内関係抽出のためのデータフォーマットの提案

水戸 拓実<sup>†,a</sup> 白井 涼子<sup>‡,b</sup> 波多野 賢治<sup>†,c</sup> 松下 光範<sup>‡,d</sup>

†同志社大学文化情報学部 ‡同志社大学大学院文化情報学研究科 ††関西大学総合情報学部

a) [bij0139@mail4.doshisha.ac.jp](mailto:bij0139@mail4.doshisha.ac.jp) b) [shirai@ilab.doshisha.ac.jp](mailto:shirai@ilab.doshisha.ac.jp)  
c) [khatano@mail.doshisha.ac.jp](mailto:khatano@mail.doshisha.ac.jp) d) [mat@res.kutc.kansai-u.ac.jp](mailto:mat@res.kutc.kansai-u.ac.jp)

**概要** 近年、デジタルコミック市場は急速に発展しており、ICT 技術を用いたデジタルコミックの活用拡大が求められる。デジタルコミックの活用例にはさまざまなものが存在するが、既存の情報編集技術を適用可能、かつ新しい研究分野の創出のために、我々は特にコミックのコンテンツデータに柔軟な情報アクセスが重要と考えている。そのためには、ill-formed なコンテンツであるデジタルコミックを形式的に取り扱えるようにするための枠組みが必要であるため、本稿ではコンテンツを構成する要素間の関係に着目した、コミックデータのための枠組みの提案を行う。

**キーワード** コミック工学, 質問応答, RDF

## 1 はじめに

タブレットやスマートフォン等、デジタル端末で読むことのできるデジタルコミック市場が急速に発展しており、その市場規模は 2011 年度で 514 億円にまで成長してきている [1]。コミックはこれまで漫画家やアニメ制作者の経験や感性に基づいて制作が進められてきたが、そのデジタル化によって ICT 技術が取り入れられたため、今後、これまでの利用形態を超え、より柔軟で利用者主導のコンテンツアクセスサービスが求められる可能性が高くなったとも言える [2]。

しかし、現在のデジタルコミックに対する情報アクセスは、単に紙媒体のコンテンツをスキャナで取り込んでいるのと大きく変わらず、デジタルコンテンツならではの利便性を十分に活用できていない。そのため、新たなサービスを産み出すためには、デジタルコミックに対する情報アクセスが、単なる書誌情報へのアクセスからコミックコンテンツそのものへ変化していく必要がある。

このような問題を改善し、書誌情報だけでなくコンテンツをも対象にした柔軟な情報アクセスを可能にするために最も必要な技術の一つは、デジタルコミックを形式的に取り扱えるようにするための枠組みの提案である。これまでも様々なタイプのコンテンツを形式的に扱えるよう、データの枠組みの提案を行う研究は行われてきた [3] が、デジタルコミックのように画像とテキストが相補的な役割を担うマルチモーダルなコンテンツではなかった。

そこで本稿では、マルチモーダルなコンテンツに対して付加情報をアノテーションとして付与することで、コンテンツにおける要素間の様々な関係を記述できるよう、

RDF [4] を用いて実現する。これにより、コンテンツの要素間の関係を記述することが可能であるため、コミックコンテンツに対しては質問応答に活用でき、前述のような利用者の要望に応えられるだけでなく、コミックに対する新しいサービスの創出につながる。

## 2 関連研究

三原らは、コミックのデータを電子的に格納するための記述フォーマットに関する研究を行っている [5]。この研究では、コミックの構成要素や書誌情報についてのメタデータスキーマを提案しているが、コミックのストーリーに関するデータは記述されないため、本稿のようにコミックコンテンツに対する質問応答を対象とするような用途には使用することができない。

また三崎らは、テキストを対象とした質問応答システムを実現するために、質問応答の回答候補を生成する際に物語の各文に記載されている動作が一つになるという、人間の物語理解の過程を基に考え出された方法を採用している [6]。

本稿の提案は、コミックコンテンツの各要素間を記述する際に文献 [6] の一知見である人間の物語理解のメカニズムを適用することで、各要素間の因果関係を手作業ではありながらも入力できる点が、これらの研究と異なる。

## 3 RDF を用いたコミックデータの枠組み

1 節で述べたように、本稿で提案する枠組みは書誌情報だけでなくコンテンツをも対象にした柔軟な情報アクセスを可能にするために、ill-formed なコンテンツであるデジタルコミックを形式的に取り扱えるようにするためのものである。前述した文献 [5] では、コミックデータの記述に OWL [7] を用いている。OWL を利用している理由は、記述内容に形式意味論を用いて語彙を

補足して提供することによって、計算機によるコンテンツ理解を実現するためである。しかし、本稿においては計算機によるコンテンツ理解ではなく、コンテンツを構成する要素間の関係に着目しているため、要素間の関係の記述が行える RDF を利用する。

コミックの書誌情報は、多くのデータ記述形式でも採用されている Dublin Core [8] を採用し、コミックのタイトルや著者などのデータを扱えるようにしている。また、コミックコンテンツの構成要素に関しては、コマを中心にコマ内に描かれるオブジェクト (含 キャラクタ)、台詞、オノマトペ、出来事などを扱えるように設計している。オブジェクトについては、特にキャラクタを表現している場合はコマ内の動きも扱い、動作の対象が存在する場合はその対象についても扱えるようキャラクタと対象間の関連を行っている。同様に出来事については、その因果関係をキャラクタの動作や台詞に関連づけることで、コミックコンテンツに対する質問応答が実現できるよう工夫をしている。

#### 4 提案手法の評価

3 節で提案した枠組みの有用性を確かめるため、コミックに対する質問応答の回答候補がどれほど抽出可能なかを評価した。

実際に用いたコミックは、「ブラックジャックによろしく」[9] の 1～3 巻である。現時点では漫画の PDF データから、本稿で提案した枠組みに機械的に変換することが困難であったため、手作業で枠組みの記述を行った。また、前述の文献 [2] を参考にコミックコンテンツに関する質問を網羅的に作成し、その質問に相応しい回答候補のデータ抽出を行った。回答データの抽出には、RDF に対する問合せ言語である SPARQL [10] を用いた。

表 1 からわかるように、提案した枠組みではコミックコンテンツの構成要素間に存在している因果関係や対象関係を表現しているため、既存手法である [5] では抽出が難しい質問でさえ、回答候補のデータ抽出がなされている。

表 1 評価結果

番号	質問	提案手法	文献 [5]
1	北三郎が登場するのは何巻の何話か？ 「手術の“ウデ”は医者が多すぎたら鈍くなるんだよ…」という台詞があるマンガはなんですか？	○	○
2	白鳥貴久の所属する医局はどこですか？	○	○
3	北三郎が歌っている歌の歌詞はなんですか？	○	×
4	2 巻で斉藤英二郎が落ち込んでいる原因はなんですか？	○	×
5	なぜ白鳥貴久は突然服を脱ぎだしたのですか？	○	×
6	研修医が様々な研修先で医者や患者と関わる中で医者という職業を知っていくマンガはなんですか？	×	×

#### 5 おわりに

本稿では、デジタルコミックに対する新しいサービスの創出のために、コミックコンテンツを構成する要素間の因果関係や対象関係を記述することが可能な枠組みの提案を行った。また、実際のコミックからコミックコンテンツを手作業で枠組みの記述を行い、質問応答タスクにおける回答候補のデータが正しく抽出できることを確認した。

今後の課題は、質問応答タスクだけでなく、より幅広くコミックコンテンツを活用するための新たなコミックコンテンツサービスを提案し、それに相応しい枠組みの再定義を行っていく予定である。また、実際のコミックを提案した枠組みで記述する際の作業負担を軽減できるよう、その半自動化の手法を提案していく予定である。

#### 謝辞

本稿では、二次利用の自由化が宣言されている、佐藤秀峰氏の「ブラックジャックによろしく」を実験に利用した。ここに記して謝意を表す。

#### 参考文献

- [1] インプレス R&D インターネットメディア総合研究所 (編) . 電子コミックビジネス調査報告書 2012. インプレス R&D, 2012.
- [2] 福田美沙紀, 白水菜々重, 松下光範. コミックを対象とした質問応答技術のための基礎検討. 人工知能学会第 2 種研究会ことば工学研究会資料, 第 40 巻, pp. 57–62, 2012.
- [3] 白井涼子, 波多野賢治. 不定型な関連性の格納が可能な和歌データベースの実装. 情報処理学会研究報告, Vol. 2013-DD-89, No. 3, pp. 1–6, 2013.
- [4] W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/rdf-concepts/>, 2004. accessed April 18, 2013.
- [5] 三原鉄也, 永森光晴, 杉本重雄. デジタルマンガにおけるストーリー構造とビジュアル構造を表すメタデータモデル. デジタル図書館, No. 41, pp. 19–26, 2011.
- [6] 三崎貴裕, 上之蘭和宏, 榎津秀次, 古宮誠一. 状況モデルに基づく質問応答システムの提案: 状態遷移テーブルを用いた状況モデル表現法. 電子情報通信学会技術研究報告, Vol. 108, No. 449, pp. 7–12, 2009.
- [7] W3C. OWL Web Ontology Language Overview. <http://www.w3.org/TR/owl-features/>, 2004. accessed April 18, 2013.
- [8] DCMI. Dublin Core Metadata Element Set, Version 1.1. <http://dublincore.org/documents/dces/>, 2012. accessed April 18, 2013.
- [9] 佐藤秀峰. 漫画 on Web. <http://mangaonweb.com/>. accessed April 18, 2013.
- [10] W3C. SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query/>, 2008. accessed April 18, 2013.