Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

RDF を利用した和歌データの管理に関する提案

白 井 涼 子[†1]　波多野 賢治[†2]

現在，和歌研究では計算機による和歌データの分析がなされており，和歌研究者の一部では新しい知識発見に寄与するシステム開発を望む声もある．和歌の原本は一冊しか存在せず，多くの人によって書き写されて普及していくことが一般的であるが，その書写過程において，誤字や表現の違いが発生し，同一和歌であっても異なる表記になっていることがある．和歌研究を進めていく上ではそれらの和歌を同一の和歌として扱う必要があるため，同一の和歌の一元的な管理を可能にすることが望まれる．

そこで，本稿では RDF による和歌データ管理手法を提案する．RDF による和歌データ管理を行うことで，和歌に関連するあらゆるデータを紐づけることが可能となり，同一の和歌を同一として扱うだけではなく，和歌単体を眺めるだけでは気づかない新たな知識発見に役立たせることが可能となる．

An RDF Representation for Classical Japanese Poems

Ryoko Shirai[†1] and Kenji Hatano[†2]

Recently, researchers tend to analyze character strings of Japanese poems using computers, and a part of them require someone to develop a system that contributes to find their new knowledge.

There is just one original copy of their anthology in the world, so that it becomes widespread thanks to transcription by a lot of people. In their transcription process, however, a literal error and a difference in their spellings may be occurred. As a result, transcriptional Japanese poem is different from its original one. If the researchers want to proceed with their research, original and transcriptional ones should be consolidated.

In this paper, we propose an approach for managing Japanese poems using RDF. It can help to bind all data related to a Japanese poem. As a result, we can turn new knowledge into a solution in the research field of Japanese poem.

†1 同志社大学大学院文化情報学研究科
Graduate School of Culture and Information Science, Doshisha University
†2 同志社大学文化情報学部
Faculty of Culture and Information Science, Doshisha University

1. はじめに

近年，和歌研究において計算機が使用される場面が増えている．その代表として， CD-ROM 版の新編国歌大観[3] や私家集大成[4] などが普及しており，キーワードや歌番号などから利用者の検索要求を満たす和歌を検索することが可能である．しかしながら，一部の研究者からは通常の検索だけでなく，知識発見に役立つ機能を有した和歌データベースを求める声もあがっている．

和歌研究の基本的なスタイルは和歌集全体を俯瞰し，研究対象としての価値が見受けられる和歌を抽出することから始まる．抽出した和歌に対して構成文字列の種類が類似した和歌[7] や，異なる和歌集に記載されている同一和歌など，何らかの関連があるとされる和歌同士を比較することや，頻出文字列を用いた比較分析[5] により和歌集ごとの傾向を測るなど，新たな知見を獲得することが和歌研究の成果といえる[6]．構成文字列の種類が類似している和歌を発見する手法を用いれば，異なる和歌集に記載された同一の和歌や有名和歌を一部改変した和歌などを発見することができるとされているが，その和歌が元は同一の和歌であったのか，それとも単に文字列の構成が類似しているだけなのかは，計算機では判断がつかないという問題点がある．このことから，和歌研究者にとっては複数の写本に写された和歌が元は同一の和歌であることをデータとして持っておき、新たな知識発見に結びつくような工夫が必要となる．

そこで，本稿では，リソースを示すことによりデータの一意性を持つことのできる Resource Description Framework (RDF) を用いた和歌データの管理手法について提案を行う．RDF を用いることにより，同一の和歌を一元的に扱うことができ，同一の和歌や付随するデータについて参照が容易になる．

2. RDF

RDF とはウェブ上でメタデータの記述を行える枠組みである．類似した機能を持つ言語として Extensible Markup Language (XML) がある．XML もメタデータを記述することが可能であるが， XML はデータ構造のみ表現可能である一方，RDF はデータ構造だけでなくリソースを用いてデータ同士の関係性を表現することが可能である．RDF は有効グラフで表されるが， XML に則った形での記述が可能で， RDF を XML 形式で表現した場

Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

合は，一般的に RDF/XML 構文と呼ばれる．

近年では，ネットワーク上のデータを参照する Linked Data というデータ管理手法もあるが，このモデルの記述にも RDF は利用されている．

2.1 基本的な RDF の構成

RDF では， Subject (主語)， Predicate (述語)， Object (目的語) の三つの要素 (トリプル) で構成される意味モデルを持つ．このとき，主語はリソース，述語はプロパティ，目的語はプロパティの値をとり，トリプルの関係は有効グラフで表すことができる．データ参照の際， RDF では Uniform Resource Identifier (URI) 参照を行う．この URI は一定の書式によってリソースを示す識別子のことであり，リソースの場所と名前によって表現され，一意に特定することが可能である．

図 1 にトリプルの有効グラフの例を示す．この例では，リソースがメディア情報学研究室，プロパティが Web サイト，プロパティの値が `http://www-ilab.doshisha.ac.jp/` となる．これは，メディア情報学研究室の Web サイトは `http://www-ilab.doshisha.ac.jp/` である，という意味を示している．つまり，リソースは説明を受けるもの，プロパティはリソースから見たプロパティの値の意味づけ，プロパティの値は実際にどういったものであるかを示している．プロパティの値にはリソースだけでなく，文字列をおくことも可能である．述語であるプロパティの値は同時に主語であるリソースにもなることが可能であるため，さらにプロパティが発生して派生する可能性がある．

図 1 トリプルの例
Fig. 1 an example for RDF triple

2.2 語　　彙

語彙とは RDF を記述するために必要な表現規則であり， RDF Schema (RDFS)[2] によって記述される．RDFS では，個々のプロパティの定義やプロパティ同士の関係を定義可能である．また，同じ性質のリソースをグループ化するクラスを定義することが可能である．このとき，クラスを表現するための基本クラスや基本プロパティといった基本語彙を提供されており，自由に語彙を設計できるようになっている．最近では，必要なクラスやプロパティをすべて最初から作成するのではなく，他の人が作成した語彙が公開されている場合には，それが利用することが多い．

一般的に， RDF を記述する際には，語彙によるクラスやプロパティ表現の異なりが発生しないように，まずは既存の語彙で利用できるものがないかを調べ，適した語彙のプロパティがない場合にのみ語彙を作成する．このとき，一つの語彙ですべてを網羅する必要はなく，複数の語彙を組み合わせることも可能である．

ここで，既存の語彙の例として，代表的な Dublin Core[1] について述べる．Dublin Core とはウェブや文書の書誌的な情報を記述するための語彙であり，基本となる 15 の要素を用いて意味を表現している．たとえば，リソースに与えられた名前を示す `title` やリソースの内容の説明を示す `description` など書式を表現する要素が含まれている．それぞれの要素が広い概念をカバーしているため，さらに詳細に示す場合には定義域や値域も設定されているため，`title` の正式タイトルの代替である `alternative`，`discription` の目次を指定する `tableOfContents`，そして要約を指定する `abstract` など，細かい指定が可能な拡張プロパティも併せて使用していく必要がある．

3. 和歌データ

和歌データとは，和歌本文に加えて和歌集に書かれている和歌に付随しているデータも含まれる．これについては 3.2 節で詳しく述べる．

ちなみに，和歌本文とは，奈良時代前後から日本で作成された定型的な韻文の歌である．標準的に五音または七音の句から構成され，代表的な和歌として知られる短歌は，五音・七音・五音・七音・七音という五つの句から構成されてる．

3.1 和　歌　集

和歌集とは，何かしらの題材や人物について和歌を編集したものであり，天皇や上皇の命により編集された勅撰和歌集と，個人が撰出した私撰和歌集がある．この編集された和歌集のオリジナルは多くの人々によって書き写されていくことにより広まり，現代に伝わってきた．その伝播過程は図 2 のような系統樹によって表現することができる．矢印は写本の書写過程を表しているが，現代ではオリジナルの和歌集が残っていることは稀であり，書き写された和歌集である写本が複数存在することが多い．

写本は書写過程での写し間違いや，所持者が故意に表現の変更を行うこともあったことから，内容が同じ写本は存在しないと言われている．また，所持者が学習のため，本文の横にメモ書きを残している場合もある．

Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

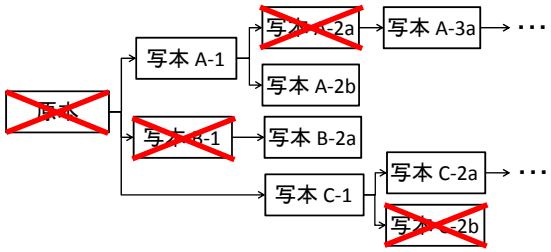

図 2 和歌集の系統樹
Fig. 2 phyletic tree of Japanese Poem

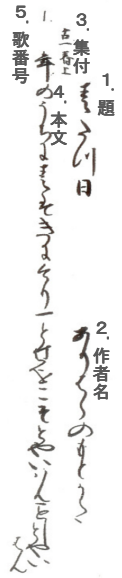

図 3 和歌データ
Fig. 3 Japanese Poem data

3.2 基本的な和歌データ

写本に書かれている和歌データの具体例を図 3 に挙げる．記述されている和歌データをまとめると以下のようになる．和歌の本文と歌番号はどの和歌データにも出現するが，題，作者名，集付は和歌によって存在するものと存在しないものがある．

(1) 題：和歌のタイトル
(2) 作者名：和歌の作成者
(3) 集付：他の和歌集の出典
(4) 本文：和歌の本文
(5) 歌番号：和歌集ごとにつけられた番号

ここで注注目するべきなのは集付である．集付は他の和歌集の出典とあるが，同一和歌が他の和歌集に出ていたことを示している．もともと和歌集は誰かが詠んだ和歌を集めているものであり，別の和歌集で用いられた和歌も記載されている場合がある．それらを，写本の書写者もしくは所持者が書き入れたものであり，他にも本文に対して誤った書写に対する訂正や注意書きを行う書入れも存在する．また，そのような書入れは認識されているものの，和歌研究者に広く普及している新編国歌大観には反映されていない．

3.3 研究で用いる和歌データ

3.2 節の和歌データを用いた和歌データの検索では新たな知識発見を行うにも限度があり，新たな要素も求められている．

通常の和歌検索ではキーワード検索や文字列検索，歌番号での検索がある．和歌検索において文字を対象とした検索の場合，漢字とかなとの違いといった表記の揺れが多く存在する．このため，本文表記の検索と併せて和歌のよみを用いての検索も行われている．しかしながら，表現の変更や写し間違いなどで本文が改変されている恐れがあるため，あるキーワードで検索した場合に同一和歌であっても表現の異なる和歌は結果に含まれないことがある．その中で書写過程における写し間違いにより表現が異なっている事例に対してには，本文に対して間違いの訂正や補足する書入れがある場合がある．表 1 の例では，本文の 1 句目には「やまかせに」とあるが，その上に「谷イ」という書入れがある．これは，その下に書かれた文字と谷と入れ替えるという指示の書入れである．この例では,「やま」の部分を「谷」に変更するという意味を持つ．このとき，書き換え後のよみのデータを生成し，本来の検索では「やま」でなければ検索結果が返されないものを,「たに」でも検索を可能にし，写し間違いであった本文でも検索可能になる．より多くの研究対象になりうる可能性を持った和歌を検索結果として返すために，書入れを修正した書き換えに対応したよみも和歌データに含める．

表 1 書入れの例
Table 1 note example

記述された本文
谷イ
やまかせに　とくるこほりの　ひまことに　うちいるなみや　春のはつ花
本文のよみ
やまかせに　とくるこほりの　ひまことに　うちいるなみや　はるのはつはな
書き換えに対応したよみ
たにかせに　とくるこほりの　ひまことに　うちいるなみや　はるのはつはな

また，3.1 節で述べた書写過程で生じる書入れは和歌集の伝播過程を知る手がかりとして，

Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

故意による表現や用語の変更は時代の風潮や言葉の流行などの時代背景を知る手がかりとして，書入れは和歌研究者にとっては重要な情報となり得る．よって，和歌データの一部として考慮する必要があり，書き込みを含めたデータを本文データと関連付けて和歌集に含まれる和歌データとする．

同様に，3.2 節で述べた集付の特徴を考慮し，同じであると考えられている異なる和歌集に存在する和歌の参照も可能にする必要があると考える．

以上のことより，利用対象となるデータは 3.2 節の，題，作者名，集付，本文，歌番号であるが，本文は実際に記述された様式と，そのよみ，さらに書入れに対応したよみを含める．

4. 提 案 手 法

本稿では 3.3 節で述べたデータを利用し， RDF を用いたデータ管理方法を提案する．

RDF を用いるメリットとして，一意にリソースを参照することが可能であることが挙げられる．今まで，和歌を同一とみなす場合，和歌集に振られている歌番号を用いた人間の目視により行ってきた．このため，関連した和歌の抽出を行う場合にも，目視で歌番号を判断して再び検索をかけると言う手間がかかっていた．人の手によって時間がかかる部分を計算機が判断できるようになると，参照しただけで付加情報の入手や付随する和歌データを簡単に抜き出し，計算機上で比較まで行える．

他に，歌番号での検索は歌番号に該当する数値が一致していることを判断しているため，歌番号が同じ和歌が結果として返されるが，計算機はそれらが同一の和歌であると認識はしていない．文字列を用いた検索と歌番号を用いた検索のどちらも，同じ和歌であっても同一和歌として管理がされていないという問題は抱えたままである．

そこで， RDF を用い，和歌集の歌番号を指定したリソースを参照することにより，同一和歌として管理可能なモデルの提案を行う．

4.1 和歌用語彙 jpoem

RDF を用いたデータの提案を行う際に，2.2 節で説明した Dublin Core と，人と人のつながりを表現可能な Friend of a Friend (FOAF) を利用するが，人物や書誌情報のみで，和歌データの表現は既存の語彙のみでは網羅しきれないという問題がある．そこで，新しく和歌に対応した語彙を定義する．和歌 （Japanese Poem） より，語彙 jpoem を準備する．jpoem はクラスに Poem のみを持ち， Poem の下に複数のプロパティを持つ構造と定義する．

プロパティの種類は表 2 に示す．和歌が持つデータは 3.2 に示されているとおり，題，作者名，集付，よみを含む和歌本文，歌番号である．このうち，作者名は Dublin Core と人間の関係性を記述する FOAF を用いて記述することが可能であり， jpoem では考慮しない．また，題も Dublin Core の使用を利用できるため， jpoem のプロパティは作成しないものとする．さらに，各プロパティの関係を図 4 に示す．

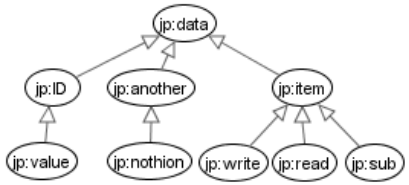

図 4 プロパティの関係図
Fig. 4 properties relation

まず，jpoem:data が和歌一種を持つものとする．ここで，一元的に管理するためのリソースは和歌集に付与されている歌番号を組み合わせたリソースをプロパティjpoem:ID で示す．さらに実際の和歌番号の数字をプロパティvalue で示す．集付のリソースはプロパティjpoem:another で示され，実際の表示は文字列としてプロパティjpoem:nothion で示す．和歌の本文の表記はプロパティjpoem:literal によって示され，そのよみはプロパティjpoem:read，書き換えに対応したよみは jpoem:sub であらわされる．さらに，その和歌の本文に関する三つのリソースはプロパティjpoem:item で示す．

表 2 jpoem:Poem のプロパティ
Table 2 property of jpoem:Poem

プロパティ	中身
data	和歌データ全般
ID	歌番号の ID
value	歌番号の数値
another	集付による参照
nothion	集付の実際の値
item	和歌本文に関するデータ
literal	和歌本文が記述されたデータ
read	和歌本文の読み方のデータ
sub	和歌の書き換えを行った後の読み方のデータ

Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

4.2 具体的な記述例

記述した例として古今和歌六畳という和歌集を用いた．そのうち，題，作者名，集付け，のある宮内庁書陵部蔵桂宮本 (桂宮本) の歌番号 1 の和歌と，本文に対する書入れのある北岡文庫蔵永青文庫本 (永青文庫本) の歌番号 5 の和歌を選択した．さらに，同一和歌であることを参照するリソースを表示するため，永青文庫本の歌番号 1 の和歌も例に挙げた．

図 5 に RDF のグラフの具体例の図を示す．これは，メタデータの記述を行えるソフトウェアの mr$^{3\star 1}$ を用いて表されている．

図 5 において，左端に位置する外向きのリンクしか持たないノードが `jpoem` のクラス `Poem` である．これは写本ごとに存在する．このとき， `dc:contributor` によって参照されているリソースは写本の所蔵を示しており，今回は桂宮本と永青文庫本のリソースを参照し， `dc:publisher` を用いて名前の文字列を示している．次に， `jpoem:data` によって参照された空白ノードはそれぞれの和歌の和歌データ示している．`jpoem:ID` で参照されるリソースはどの写本のどの歌番号かを示しており，今回は古今和歌六畳の 1 番と 5 番に対応しており， `jpoem:value` によって歌番号の文字列を持つ．この `jpoem:ID` で示されたリソースが特定の和歌集に付与された歌番号のデータ，`rokujo.001` と `rokujo.0005` をそれぞれ参照しており，これを用いることにより同一和歌であることを示し，一元的に管理を行うことが可能である．このリソースと和歌を抜き出した部分を図 6 に示す．

`dc:title` では和歌の題を， `dc:creator` では作者を参照している．今回は歌番号 5 の和歌には存在しなかったが，歌番号 1 の和歌にどちらも存在した．桂宮本と永青文庫本があるが，題はどちらも「春たつ日」である．しかし，作者の表記が異なり桂宮本では「ありはらのもとかた」，永青文庫本では「あり原のもとかた」であった．集付はほかの和歌集の出典として,「古一春上」が書き込まれていた．これは古今和歌集第一巻 (春の上巻) を示している．この古今和歌集の和歌と同一和歌であることを示すためのリソースとして，`jpoem:ID` にて `kokin.0001` を参照している．古今和歌集の歌番号は 1 であることもわかる．和歌の本文については，永青文庫本の歌番号 5 の和歌に着目する．和歌の参照の前に，複数の値をまとめるクラスとして `Seq` クラスを用いる．これはシーケンスコンテナのためのクラスであり，参照するコンテナのそれぞれに順序を付与し，その順番が重要なことを示している．この順序は句の順番に付与した．その後，和歌の本文の表記を `literal`，そのよみを `read`，書入れに対応したよみを `sub` によってそれぞれの句を参照可能にした．

⋆1 `http://mr3.sourceforge.net/ja/`

5. おわりに

本稿では RDFS を用いて和歌データのモデルの記述ができた．今後の課題として，この RDFS に基づいて RDF データベースを構築し，それに和歌データの格納を行う．

また，さらに発展させる形として， wikipedia のようにデータを持つものから，作者名からリンクを結び， Linked Data 化してより多くの関連したデータを集約することにより，和歌研究者の知識発見という要望に対応できるのではないかと考えている．

謝辞 本研究の一部は日本学術振興会科学研究費補助金 若手研究 (B)(課題番号:22700248) と同志社大学大学院文化情報学研究科研究推進補助金によるものである．ここに記して謝意を表す．

参考文献

1) ISO: The Dublin Core metadata element set. `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52142`.
2) W3C: RDF Schema. `http://www.w3.org/TR/rdf-schema/`.
3) 「新編国歌大観」編集委員会（編）：CD-ROM 版 新編国歌大観，角川学芸出版 (1996).
4) 『私家集大成』CD 化委員会（編）：新編私家集大成 CD-ROM 版，エムワイ企画 (2008).
5) 斉藤康彦：頻出文字列に基づく古今和歌集と新古今和歌集の比較分析の試み，情報処理学会研究報告. 人文科学とコンピュータ研究会報告，Vol.2009-CH-81, No.4, pp.25–32 (2009).
6) 福田智子：古典和歌研究における計算機科学の有用性，文化情報学入門（村上征勝，編），勉誠出版，chapter3 (2006).
7) 竹田正幸，福田智子：古典和歌からの知識発見 -モビルスーツを着た国文学者-，情報処理，Vol.43, No.9, pp.941 – 949 (2002).

Vol.2012-IFAT-106 No.2
Vol.2012-DD-85 No.2
2012/3/26

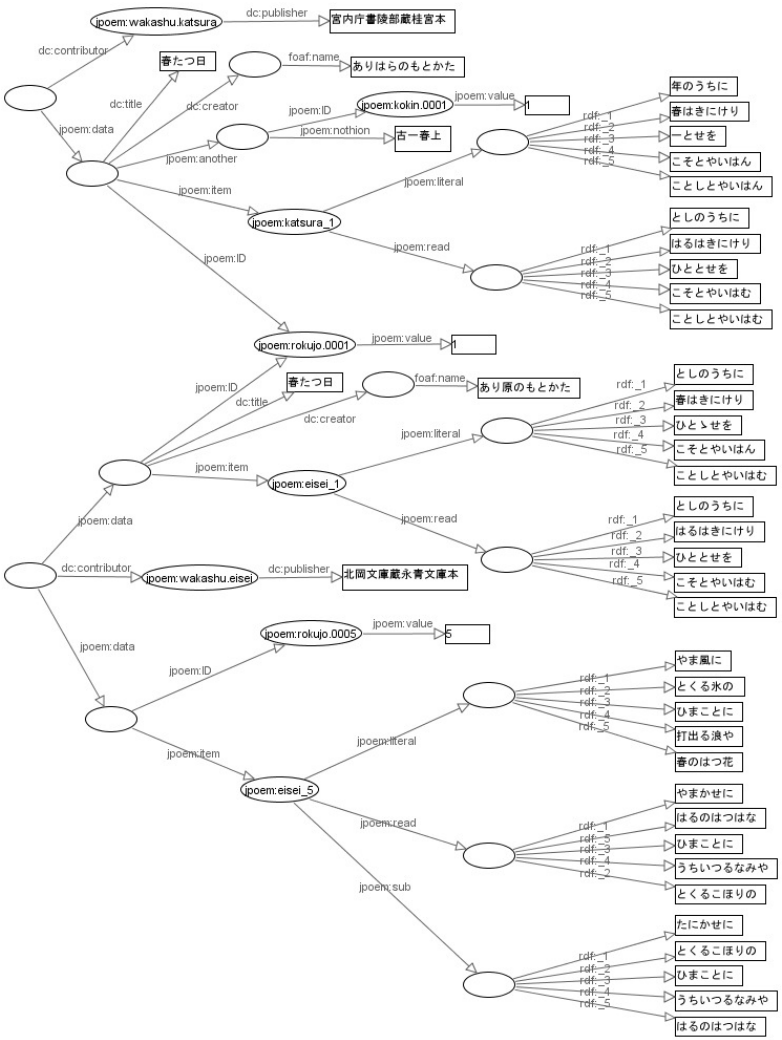

図 5 記述された RDF のグラフの例
Fig. 5 example RDF graph

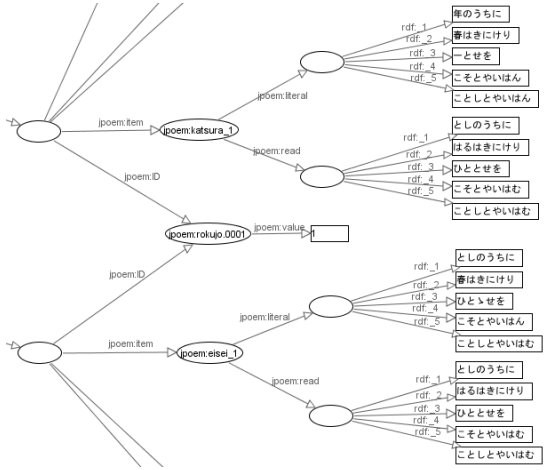

図 6 和歌を統括するリソース
Fig. 6 waka resource