仏教思想の概念体系の記述手法としての TEI マークアップの現状と課題

左藤 仁宏（東京大学大学院）
渡邉 要一郎（東京大学）
永崎 研宣（人文情報学研究所）
下田正弘（東京大学）

概要：近年，テキストに現れるさまざまな関係を記述する手法として，RDF (Resource Description Framework) に基づく Linked Data を用いる取り組みが広まりつつある．2020 年 8 月，そうした流れを受け TEI P5 ガイドラインでは，version 4.1.0 において，model.standOffPart を含む <standOff> エレメントを追加した．本稿では，Linked Data を典籍の本文と共存させつつ配布するこの手法について，仏教思想概念の整理がなされる典籍，親鸞著『愚禿鈔』（13 世紀成立）を事例として検討する．そして，テキストに現れる諸概念の関係を記述する方式の分類を提案し，現状で明らかになっているこの記述手法の可能性と課題について報告する．
キーワード：TEI, 知識グラフ, standOff, 日本仏教, 漢文, 愚禿鈔

Current Status and Issues of TEI Markup as a Method for Describing the Conceptual System of Buddhist Thought

Yoshihiro Sato (The University of Tokyo Graduate School)
Yoichiro Watanabe (The University of Tokyo)
Kiyonori Nagasaki (International Institute for Digital Humanities)
Masahiro Shimoda (The University of Tokyo)

Abstract: In recent years, there has been widespread use of Linked Data based on RDF (Resource Description Framework), a method for describing various relationships that appear in text, has become widespread. In August 2020, in response to this trend, the TEI P5 Guidelines version 4.1.0 added a <standOff> element containing model.standOffPart. In this paper, we examine this method of distributing Linked Data in a way that allows it to coexist with the body of the source text, using as a case study Shinran's Gutokushō (13th Century), a text in which the concepts of Buddhist thought are organized. We will then categorize the methods used to describe the relationships between the concepts that appear in the text, and report on the possibilities and challenges of this method as currently understood.
Keywords: TEI, Knowledge Graphs, standOff, Japanese Buddhism, Chinese writing, Gutokushō

１．まえがき

世界中の数多の思想体系においてと同様に，仏教においてもまた，長い時間をかけた思索が重層的に体系化され，その体系自体を記述する典籍も様々に執筆され伝承されてきた．

一方で，そうした思想的な概念体系も含む様々な関係記述の手法として RDF (Resource Description Framework) に基づく Linked Data を用いる手法がセマンティック Web を中心として広く普及しつつあり，テキストに含まれる様々な関係を Linked Data で記述しようとする取り組みも広まりつつある．そうした流れを受けて，TEI P5 ガイドラインでは，version 4.1.0 において，model.standOffPart を含む <standOff> エレメントを追加した．このエレメントの説明は以下のようになっており，

> Functions as a container element for linked data, contextual information, and stand-off annotations embedded in a TEI document. [1]

Linked Data のみならず，テキストから独立した様々な要素をも柔軟に取り込むことができるようになっている．これにより，Linked Data を典籍の本文と共存させつつ配布する手法が国際的

に提供された形となった．

２．これまでの関連する状況

典籍に基づく研究においては，ある情報がいかにテキストから離れて抽象化したとしても，テキストにおいてその根拠となる箇所を参照できることの重要性が失われることはない．そして、テキストにまつわるすべての情報をひと所に提供することは不可能だとしても，それら情報の根拠となる箇所を参照できる状態にすることは，研究成果の検証可能性の担保やその継承と発展において不可欠なものである．換言すれば，根拠となるテキストの参照が困難な情報は，それがもたらし得る可能性の多くを失ってしまっていることになる．

しかし，Linked Data のようなグラフによる知識記述は，グラフの形式に落とし込めないデータは扱うことができず，たとえばテキストに含まれる部分的な要素を対象とした記述を行うには何らかの工夫を必要とする．典型的な方法としては，なんらかの特別な仕組みを作成して URL でテキストを部分参照できるようにするか，あるいはテキスト中に ID 付きのアンカーを埋め込んでこれを参照する方法等が考えられる．

前者に関しては西洋古典学における CTS (Canonical Text Services) [4]や，漢文仏典における SAT 大蔵経データベースの 2012 年版以降[5]などが例にあげられよう．これらのデータベースでは Web を介して URN や URL でテキストを部分参照できるようにする仕組みが実装されており，これらを用いることで，そのデータベースが取り扱うテキストに関してはテキストの部分参照が可能となっている．しかしながら，この仕組みでは URN/URL 等と任意の箇所を対応づけるための何らかのプログラムが必要であり，さらに，現時点では，歴史的に長い時間をかけて整備され充実した目録が提供されているテキストでしか実現できておらず，目録が十分に作成されていないテキストデータでは実装することが難しい．したがって，汎用性を確保するという観点では十分とは言えない段階である．

一方，後者の方向性では，TEI ガイドラインを利用することである程度対応可能ではある．たとえば，本文中の任意の箇所に人名<persName>や地名<placeName>等の固有表現のタグや，あるいは参照文字列<rs>や任意の句<seg>等のタグを用いてマークアップした上で，それぞれに xml:id を付与して，Linked Data の対象となる文字列を他所から参照できるようにしておく．その上で，TEI 以外のスキーマを記述して TEI 文書に内包するための xenoData エレメント[6]にそれらを対照とした RDF/XML を記述するというのが一つの方法ではある．しかしながら xenoData は、どんなスキーマでも記述してよいことになっており，記述や処理を共通化して利便性を高めるという観点からは十分なものではない．すなわち，TEI ガイドラインのそれまでの枠組みにおいては本文の内容に対応させる形で Linked Data を適切に記述することは，不可能ではないものの，容易なことではなかったと言える．

TEI ガイドラインにおいて<standOff>が導入されたことは，上記のような事情により記述が困難であったものを同一の電子文書の中で実現できる仕組みが提供されたという点で画期的であった．

このような技術は，思想概念を整理する内容を持つ資料から，そこに記される概念の知識グラフを作成するにあたり有用である．そして，そのための TEI マークアップの方法を整備していくことは，例えば日本仏教文献に複数存在する，思想概念を整理するメモ書きのようなテキストをマークアップしていくに際して有益であろう．

そこで本稿では，親鸞によって著された『愚禿鈔』（13 世紀成立）という書物を事例として，思想概念の整理を主題とするような他の文献への適用を目論みながら，この記述手法の可能性と課題について検討したい．

３．『愚禿鈔』の概念体系における，諸概念の関係の種類

親鸞によって著された『愚禿鈔』[1]は，上巻・下巻の二巻からなり，上巻では仏教各宗の教義理論を分類し，下巻では特に信心についての教義を論じる書物である．その著述形式は独特で，特に上巻においては，仏教思想体系に含まれる概念を整理するための，メモ書きのような体裁を有している．下図 1 に示すように，写本及び刊本[2]では，仏教思想概念を列挙するなどして，それらの概念に対して改行とインデント下げ，割注を多用することで，それぞれの概念同士の上下関係を示すという形式を用いている．

1) 本発表で使用したマークアップ済み電子テキスト，および可視化結果は [3] に公開中である．

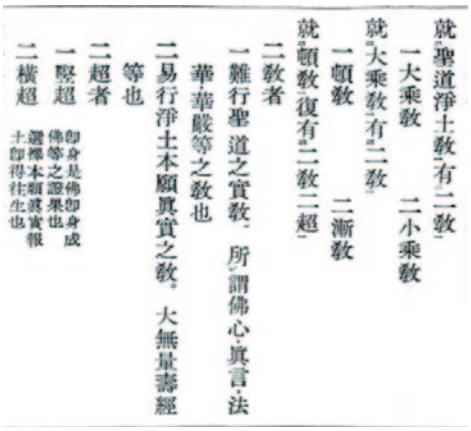

就聖道浄土教有二教
一大乗教　二小乗教
就大乗教有二教
一頓教　二漸教
就頓教復有二教二超
二教者
一難行聖道之實教、所謂佛心・眞言・法華・華嚴等之教也
二易行浄土本願眞實之教、大無量壽經等也
二超者
一竪超　卽身是佛卽身成佛等之證果也
二横超　選擇本願眞實報土卽得往生也

図１『愚禿鈔』上巻 本文

『愚禿鈔』は全体としてこのような形式で著されており，概念の樹形図を示唆する内容を有しているという点において，本文献は<standOff>の知識グラフ作成を目論んだTEIマークアップの試行に適当であると思われる．また，その内容から概念の樹形図を示すような文献は，日本仏教の分野に複数存在し，それら他文献にも適用可能なマークアップ方式を模索するという点においても，『愚禿鈔』における試行は有益である．

本項では，『愚禿鈔』が記述しようとする概念体系を<standOff>の知識グラフとして表現するために必要な，文献に現れる概念同士の関係の種類について記したい．そして次項にて，それぞれの種類の関係に対応した TEI マークアップの方法，知識グラフへの反映のさせ方について論じる．

A. 親子関係（hasParent）

例えば，『愚禿鈔』本文中では「<聖道浄土教>について二教あり，一つには<大乗教>，二つには<小乗教>なり」（図１ 1-2 行目）と記される例がある．ここでは，<聖道浄土教>という名詞概念が親，<大乗教>，<小乗教>という名詞概念がそれぞれ子に相当し，これらの概念が親子関係をなしていることは明らかである．

B. 同一関係（sameAs）

また『愚禿鈔』本文では上の記述に続いて，「<大乗教>について二教あり，一つには<頓教>，二つには<漸教>なり」（図１ 3-4 行目）と記される．ここで現れる<大乗教>という名詞概念は，前文に出現した<大乗教>（図１ 1-2 行目）と同義語であることから，両者が同一関係をなしていることが認められる．この同一関係は，後述する E. 類似関係よりも強い，完全な同義語に対して認められる．

なお，ここでも二度目の<大乗教>の子概念として<頓教>と<漸教>の二つが置かれており，前述の親子関係が示されていることがわかる．

C. 説明関係（explanationOf）

ある概念に対して，別の名詞概念，形容詞概念，形容句概念が補語として説明を加えていると見做せる場合に，そこに説明関係があると見做すことができる．以下でその実例を紹介しよう．

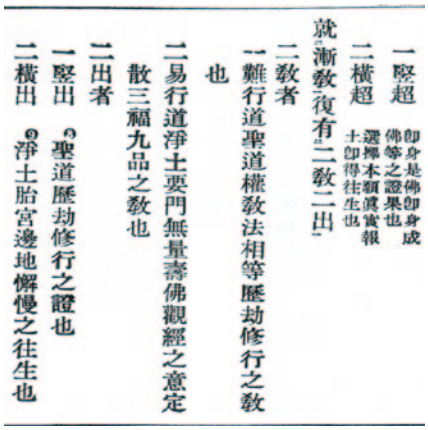
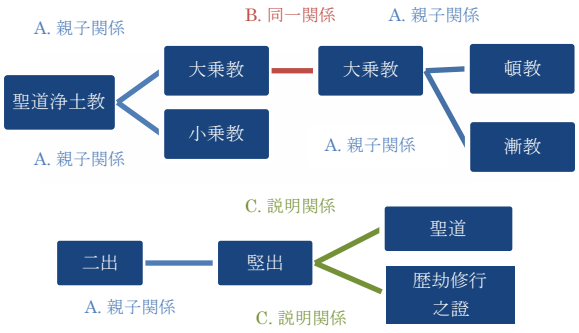

一竪超　卽身是佛卽身成佛等之證果也
二横超　選擇本願眞實報土卽得往生也
就漸教復有二教二出
二教者
一難行道聖道權教法相等歷劫修行之教也
二易行道浄土要門無量壽佛觀經之意定散三福九品之教也
二出者
一竪出　聖道歷劫修行之證也
二横出　浄土胎宮邊地懈慢之往生也

図２『愚禿鈔』上巻 本文

上図２は，図１で示した本文の続きである．例えばここでは「<二出>とは，一つには<竪出>，<聖道>，<歴劫修行之證>なり」（図２ 9-10 行目）と記される．<二出>の子概念として<竪出>という概念が置かれ，この<竪出>が<聖道>（自力で悟りを得ようとする道）であり，<歴劫修行之證>（量り知れない時を経た修行によって悟ること）であると述べられている．このようなとき，<聖道>，<歴劫修行之證>の二つの概念は独立した名詞概念というよりも，<竪出>に対する説明の関係に置かれるノードとして理解される．

以上の，A. 親子関係，B. 同一関係，C. 説明関係の三種があれば，『愚禿鈔』に現れる概念をマークアップする基本とすることができる．これら三種の関係を踏まえて，これまで述べてきた実例についての概念図を示せば，以下のようになるだろう．



さらにこれらに加えて，以下のような事例に対応するために，兄弟関係，及び類似関係について考えたい．

D. 兄弟関係（hasSibling）

本文中には，ある概念が別の概念と同階層として並列される場合があり，これを兄弟関係として認めたい．

図３『愚禿鈔』上巻 本文

例えば，上図３には「<大經>には<選擇>に三種あり，（中略）<觀經>には<選擇>に二種あり，（中略）<小經>には<勸信>に二つ，<證成>に二つ<護念>に二つ，<讃嘆>に二つ，<難易>に二つあり」と記されている．本文中には，<大經><觀經><小經>の親に相当するような概念は現れていないものの，一般に浄土教の文脈においては大經・觀經・小經は『無量寿経』『観無量寿経』『阿弥陀経』の略称として扱われ、これらが浄土三部経という、最も尊崇されるべき三つの経典として取り扱われることからも，これらの三つの概念が同じ階層で並列されていることは明白である．このようなとき，<大經><觀經><小經>は兄弟関係にあると見做せるだろう．

E. 類似関係（similarWith）

また，前項 B. 同一関係で論じた場合と違い，ある語が固定的な名詞概念とは言えない，定義がぶれやすい概念であり，かつその語と少なくとも表面上は同一の語が別の文脈で文中に出現する場合には，それらを B. 同一関係よりも緩やかな，類似関係として見做した方がいい場合がある．

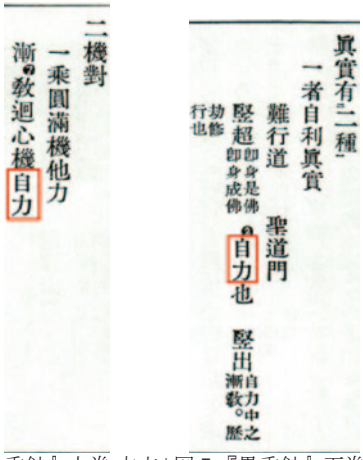


図４『愚禿鈔』上巻 本文 | 図５『愚禿鈔』下巻 本文

例えば，「<漸教迴心機>は<自力>なり」（図４ 3 行目）と上巻で記されたあと，文脈が変わった下巻において「<竪超>は<自力>なり」（図５ 4 行目）と記される場合，両者の<自力>はともに直前の名詞概念に対する補語，説明のための概念になっているが，これらを同じ語であるからといって完全な同一概念と見做していいのかどうか，判断するためには高度な文脈を読解する必要があり，作業者にとっては判然としない場合がある．しかし無関係の語であるとは見做しがたい．こういった場合の方策として，ここでは両者の<自力>を類似関係として見做したい．これによって，同形の語を全て同義語と見做した場合に発生するような，知識グラフの煩雑化を緩和する狙いもある．

以上の，A. 親子関係（hasParent），B. 同一関係（sameAs），C. 説明関係（explanationOf）の基本的な三種類の関係に加えて，D. 兄弟関係（hasSibling），E. 類似関係（similarWith）の合計五種類で，『愚禿鈔』の概念体系を記述することとしたい．

４．諸概念の関係の記述方式

具体的なマークアップの手続きとしては，本文中に出現する各概念に，<seg> タグで必要な情報を手作業で付与する．そして文中に現れる各単語に付与した <seg> には，xml:id 属性を付与する．また，同じ概念，同じ単語が複数回出現している場合には，例えば xml:id="三福", xml:id="三福 1" というように，単語の末尾に数字等を付与し，これらを一度違うものと見做したあと，後述する B’ の方法を用いてこれらを同一視する．これによって，単語が出現するテキスト上の位置と，これらの語が出現した個別具体の文脈ごとにおける他の概念との連関を保存しつつ，テキスト全体における総合な各概念間どうしの関

連を記述することができる．

また，そうしてマークアップされた電子テキストから，自動的に standOff.graph 内に <node> と <arc> を作成する．このように生成された <graph> は，容易に turtle 形式等に書きなおすことができ，概念間の連関を可視化することも可能である．

以下に，前項で分類した A から E までの諸概念の関係の種類に対応した，具体的なマークアップの手順を記したい．

A’. 親子関係（hasParent）

まず，『愚禿鈔』全体にとっての root 概念である <聖道淨土教> という語が本文中に最初に出現したとき，これを

<seg xml:id="聖道淨土教">聖道淨土教</seg>

と，マークアップする．その後，この概念の下位階層に属する概念である<大乗教>という語が現れた箇所を

<seg xml:id="一大乗教" corresp="#聖道淨土教" type="hasParent">一大乗教</seg>

とマークアップする．このとき，xml:id には本文に出ている通りの形を登録し，@corresp で親に相当する概念を参照する．

このような記述は，<graph> 内では

<node corresp="#聖道淨土教"> <node corresp="#一大乗教"> <arc to="#聖道淨土教" from="#一大乗教" ana= "hasParent"/>

と記述されるように変換を行う．

以下，B’ から D’ の項目についても，<graph> 内の記述に関しては @ana を適宜変更するだけであるから省略する．

B’. 同一関係（sameAs）

続いて，この直後に現れる「大乗教」は，前述した通り「一大乗教」と同一の概念なので，この二つの語を同一視するため，以下のように記述する．

<seg xml:id="大乗教" corresp="#一大乗教" type="sameAs">大乗教</seg>

C’. 説明関係（explanationOf）

さらに，前項 C. 説明関係における実例をマークアップするときには，以下のようになる．<聖道>という語が，<竪出>という語を説明しているので，

<seg xml:id="聖道" corresp="#竪出" type="explanationOf">聖道</seg>

と，このように記述する．

D’. 兄弟関係（hasSibling）

「大經」が「觀經」「小經」と兄弟関係にある．この際の兄弟関係のマークアップは最小限でよい．例えば「大徑」が出現する箇所に対して

<seg xml:id="大經" type="hasSibling" corresp="#觀經 #小經">大經</seg>

とマークアップを施し，テキスト中に出現する「觀經」「小經」に対しては <seg xml:id="觀經">觀經</seg>, <seg xml:id="小經">小經</seg> とだけ記述する．兄弟関係は双方向的であるので，後ほど自動的に三者間に兄弟関係のネットワークが <graph> 内に生成されるようにする．

E’. 類似関係（similarWith）

まず，最初に現れる「自力」は「漸教迴心機」に対する説明概念になっているので，これを

<seg xml:id="自力" corresp="#漸教迴心機" type="explanationOf">自力</seg>

とマークアップする．次に，再び文中に現れる「自力」は，「竪超」に対して説明関係にあり，かつ前出の「自力」に対しては類似関係にあるという，二つの関係を有しているため，以下のように二重にマークアップすることとする．

<seg type="explanationOf" xml:id="自力1" corresp="#竪超"><seg type="similarWith" corresp="#自力">自力</seg></seg>

５．応用可能性と課題

本稿で提案された記述方式は、他の日本仏教文献においても適用可能なものである．例えば，本稿で取り扱った『愚禿鈔』も収録されている漢文仏典の一大叢書である『大正新脩大藏経』には，以下のような体裁を持った文献も含まれている．

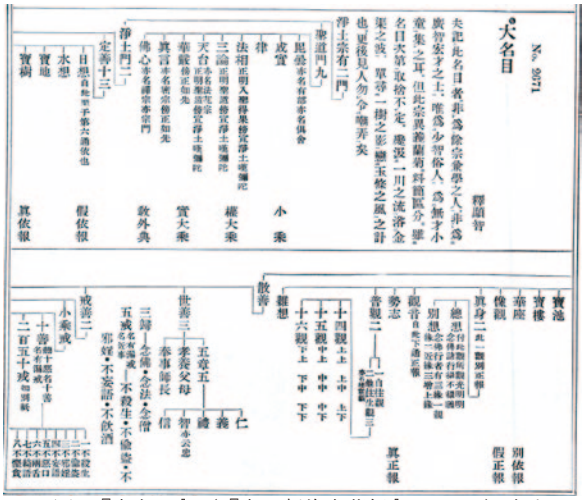


図 6『大名目』（『大正新脩大藏経』No.2671）本文

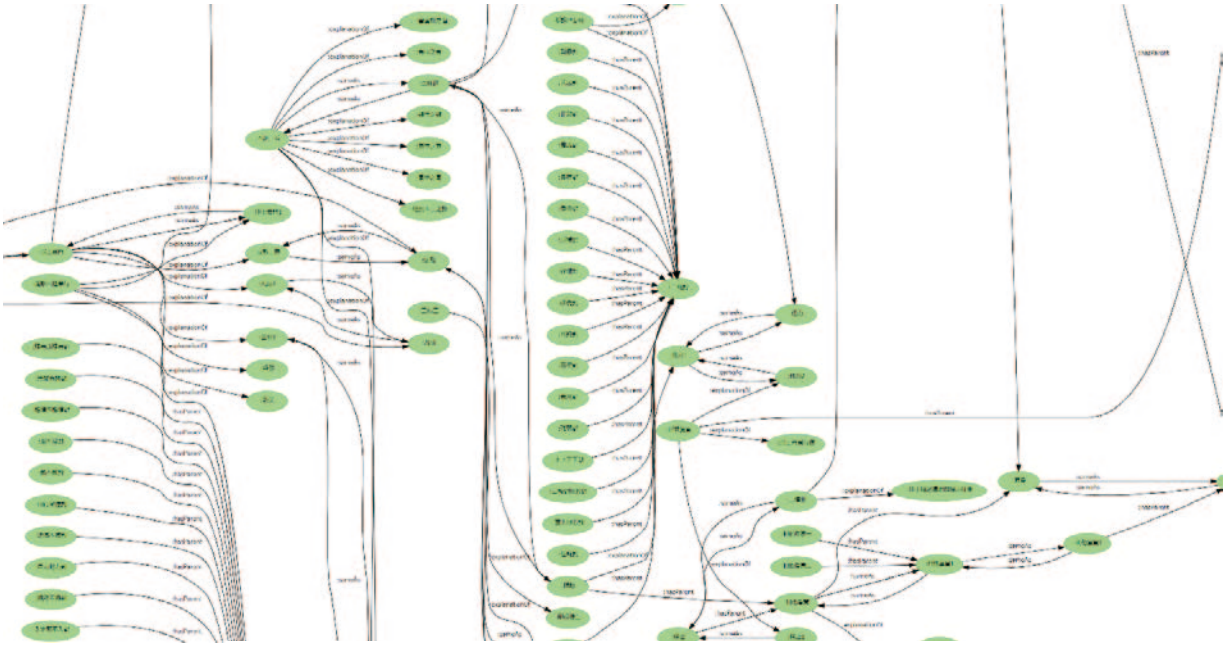


図 7〈standOff〉から作成された知識グラフの可視化の一部[7]

上図６のような，知識グラフを前提としたような文献をマークアップし、より有用なテキストを利用者に提供するにあたっても，本稿で示した記述方式が適用可能である．

また，本研究には以下のような課題も残されている．これまで述べてきたようにマークアップされたテキストから <graph> の<node> を作成するとき，例えば上述の B’ の例では@xml:id は異なるが同じノードとして扱うべきものである．この場合，そのまま<arc>に変換するのみでなく，type="sameAs"を参照しつつ一つのノードとなるように処理する必要がある．

また，例えば D’ では「大經」「觀經」「小經」などの兄弟概念のさらに親概念として，文献中には見出されない「浄土経典」などの概念を設定しておくことが，より知識グラフの有用性を高めるであろう．これについては自動処理が困難であり，<node>と<arc>に手動で入力せざるを得ないのが現状である．

TEI における<standOff>の活用には他にも様々な課題があり得るが，典拠に基づく知識グラフ記述の手法として，今後も検討を続けていきたい．

なお『愚禿抄』に出現した概念の可視化の一例を図 7 に挙げておく．

参考文献

[1] “TEI P5 Guidelines, 16.10 The <standOff> Container”. https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SASOstdf, (参照 2021-8-29).

[2]『大正新脩大蔵経』大正一切経刊行会, No.2648, Vol.83, pp.647a-654a.

[3] “愚禿鈔の RDF”. https://github.com/wyoichiro1125/gutokusyou_rdf,(参照 2021-11-1).

[4] The Canonical Text Services URN specification, version 2.0.rc.1 http://cite-architecture.github.io/ctsurn_spec/ (参照 2021-11-1)

[5] 永崎研宣他, 大蔵経における多言語対訳コーパスの構築, じんもんこん 2009 論文集 2009(16), pp. 129-134, 2009-12-11.

[6] <xenoData> https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-xenoData.html (参照 2021-11-1)

[7]https://github.com/wyoichiro1125/gutokusyou_rdf/blob/main/graph-draw_2648.svg (参照 2021-11-1). 作成にあたり https://www.kanzaki.com/works/2009/pub/graph-draw (参照 2021-11-1) を使用した．