The University of Osaka
Institutional Knowledge Archive

Title	人文学研究者必見! テキストデータとTEIで描く新たな研究ビジョン
Author(s)	吉賀, 夏子; 田畑, 智司; 甲斐, 尚人 他
Citation	
Version Type	VoR
URL	https://doi.org/10.18910/101978
rights	This content is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

TEIタグの使い方

タグとは何？

1. <xxxxx>文章や文字の塊</xxxxx>

<u>エレメント(要素)</u>＝タグに囲まれた文字の断片

開始タグ　　終了タグ

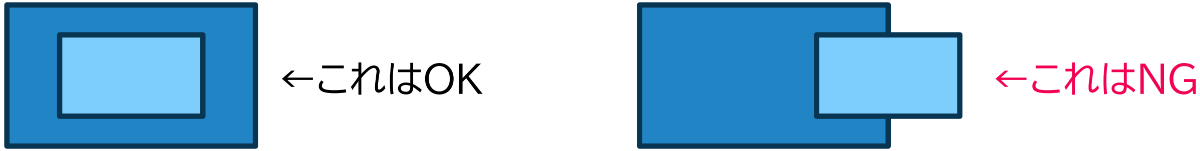
2. <xxxxx />

空白タグ(空要素)

このタグは、テキストがなくても入力した位置に<u>何らかの意味を持たせることができる</u>

タグ付けのルール

1. 開始タグと終了タグで文字の塊を囲むときは、必ずセットで使用する
2. 開始タグと終了タグのセットは入れ子にしても良い
3. 入れ子同士はオーバーラップできない



TEI形式のデータの外観

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- ... -->
 </teiHeader>
 <text>
  <front>
<!-- front matter of copy text, if any, goes here -->
  </front>
  <body>
<!-- body of copy text goes here -->
  </body>
  <back>
<!-- back matter of copy text, if any, goes here -->
  </back>
 </text>
</TEI>
```

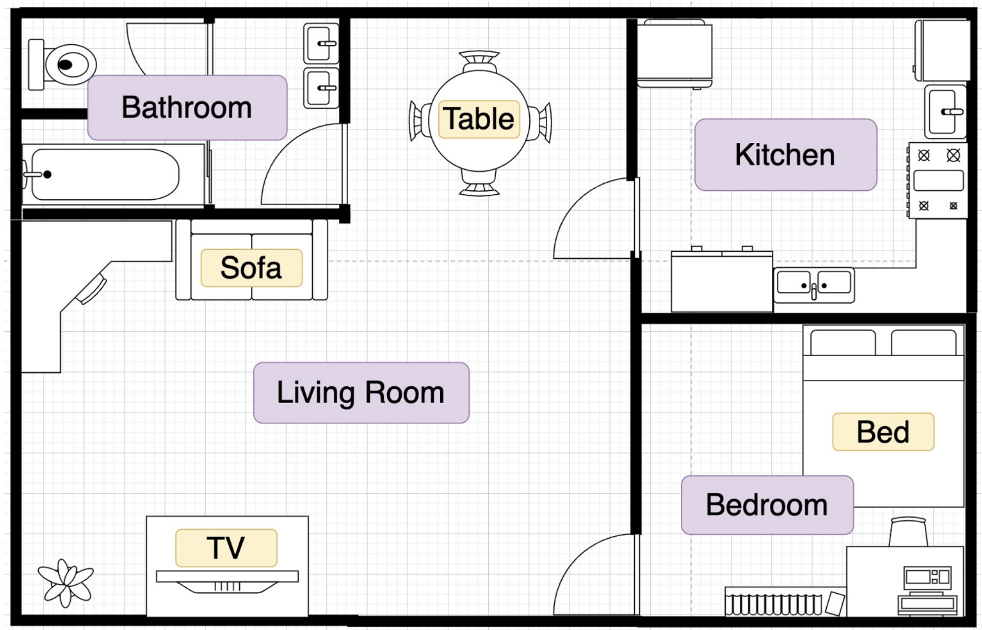
<TEI>タグ（ルート要素）

<TEI>

ルート要素

（一番最初に<u>必ず使うタグ</u>）

<TEI>



</TEI>

<teiHeader>タグ

<teiHeader>

<TEI>の中に置かれる必須の要素

テキストの周辺情報(タイトル、著者名、発行日など)を管理する部分

<TEI>

<teiHeader>

本文の概要や補足

</teiHeader>

<text>

本文

</text>

</TEI>

<text>タグ（本文部分）

<text>タグ

実際のテキストの内容
章、段落、見出しなどを整理する

<TEI>

<teiHeader>

本文の概要や補足

</teiHeader>

<text>

本文

</text>

</TEI>

TEI形式のデータの外観

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- ... -->
 </teiHeader>
 <text>
  <front>
<!-- front matter of copy text, if any, goes here -->
  </front>
  <body>
<!-- body of copy text goes here -->
  </body>
  <back>
<!-- back matter of copy text, if any, goes here -->
  </back>
 </text>
</TEI>
```

論文ファイルを使った簡単な例

（書誌情報）
題名　私の論文
著者　山田太郎
出版者　阪大出版
出版年　2024

（本文テキスト）
私の論文
山田太郎

序論
これは私の論文の序論です。
…

<TEI>
<teiHeader>
（出所＝書誌情報、目録、メタデータ）
</teiHeader>
<text>
（本文テキスト）
</text>
</TEI>

タグの構造

（書誌情報）
題名　私の論文
著者　山田太郎
出版者　阪大出版
出版年　2024

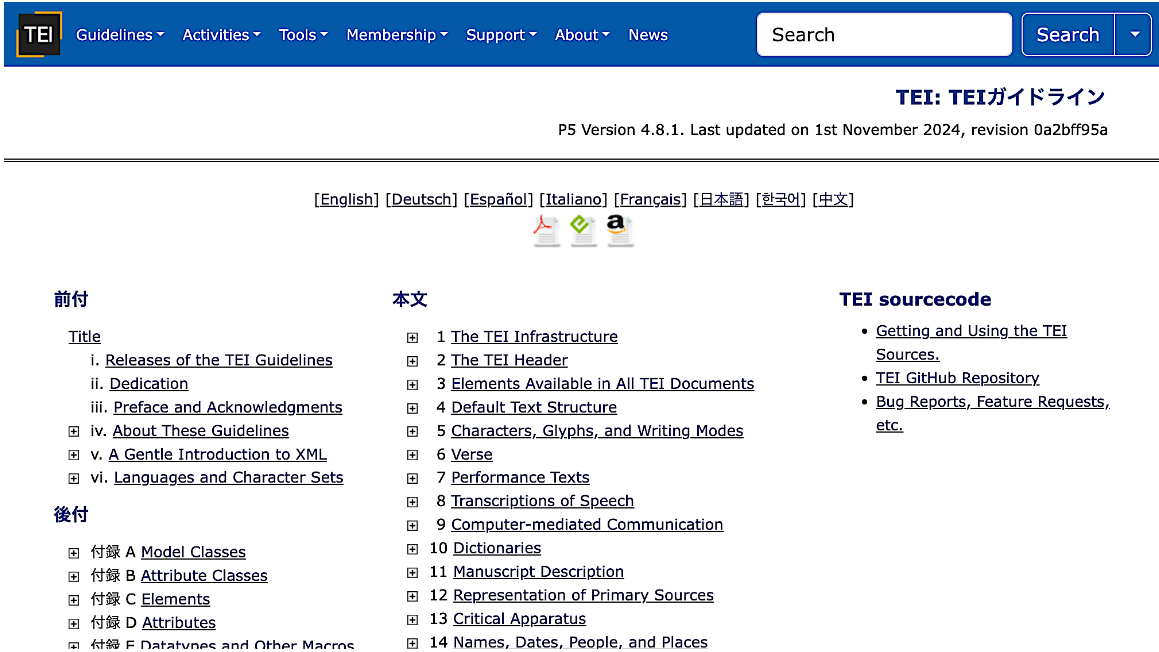
（本文テキスト）
私の論文
山田太郎

序論
これは私の論文の序論です。
…

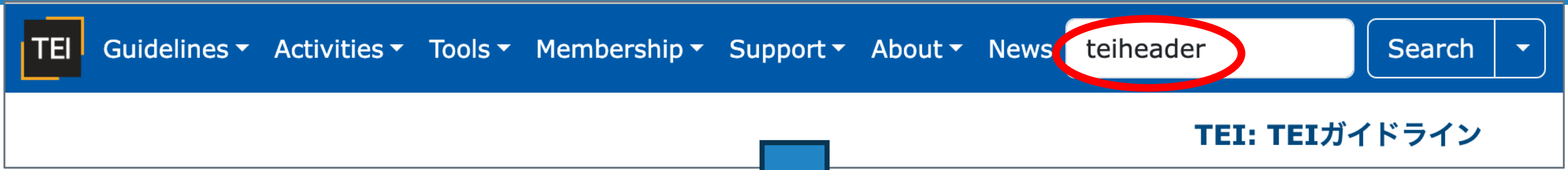
```
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>私の研究論文</title>
        <author>山田太郎</author>
      </titleStmt>
      <publicationStmt>
        <publisher>阪大出版</publisher>
        <date>2024</date>
      </publicationStmt>
    </fileDesc>
  </teiHeader>
  <text>
   （本文テキスト）
  </text>
</TEI>
```

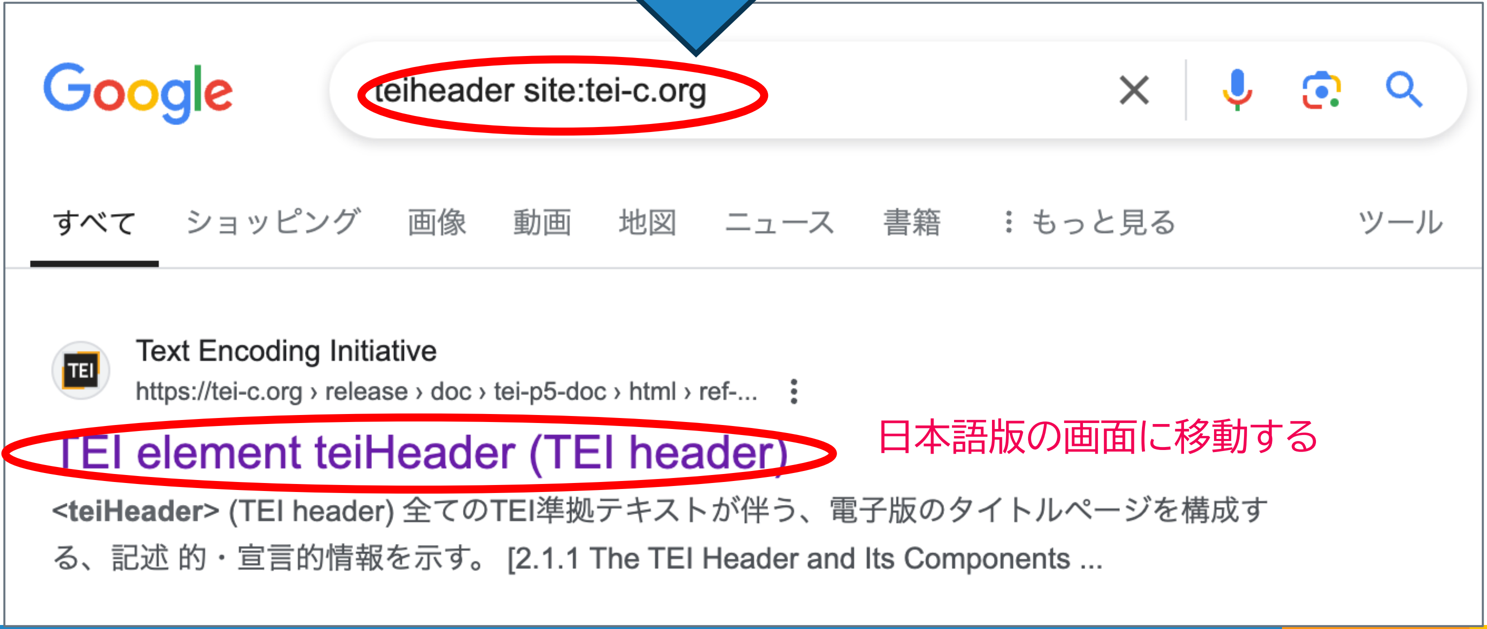
ガイドラインの使用方法(1)

「TEIガイドライン」とブラウザで検索すると、右の画面のような「The TEI Guidelines」(Current Guidelines)というページが見つかります



ガイドラインの使用方法(2)





日本語版の画面に移動する

ガイドラインの使用方法(3)

<teiHeader>

ホーム
C 要素

<teiHeader> (TEI header) 全てのTEI準拠テキストが伴う、電子版のタイトルページを構成する、記述 的・宣言的情報を示す。 [2.1.1 The TEI Header and Its Components 16.1 Varieties of Composite Text]	
モジュール	header — The TEI Header
属性	• att.global: @xml:id, @n, @xml:lang, @xml:base, @xml:space ◦ att.global.rendition: @rend, @style, @rendition ◦ att.global.linking: @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select ◦ att.global.analytic: @ana ◦ att.global.facs: @facs ◦ att.global.change: @change ◦ att.global.responsibility: @cert, @resp ◦ att.global.source: @source
上位	**core:** teiCorpus **textstructure:** TEI
下位	**header:** encodingDesc fileDesc profileDesc revisionDesc xenoData
解説	当該要素は、TEI文書中で無条件に必要となる要素の1つである。

例	<pre><teiHeader>
	 <fileDesc>
	  <titleStmt>
	   <title>Shakespeare: the first folio (1623) in electronic form</title></pre>

ガイドラインの使用方法(4)

下位	**header:** encodingDesc fileDesc profileDesc revisionDesc xenoData
解説	当該要素は、TEI文書中で無条件に必要となる要素の1つである。
例	

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>Shakespeare: the first folio (1623) in electronic form</title>
   <author>Shakespeare, William (1564–1616)</author>
   <respStmt>
    <resp>Originally prepared by</resp>
    <name>Trevor Howard-Hill</name>
   </respStmt>
   <respStmt>
    <resp>Revised and edited by</resp>
    <name>Christine Avern-Carr</name>
   </respStmt>
  </titleStmt>
  <publicationStmt>
   <distributor>Oxford Text Archive</distributor>
   <address>
    <addrLine>13 Banbury Road, Oxford OX2 6NN, UK</addrLine>
   </address>
   <idno type="OTA">119</idno>
   <availability>
    <p>Freely available on a non-commercial basis.</p>
   </availability>
   <date when="1968">1968</date>
  </publicationStmt>
  <sourceDesc>
   <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton Facsimile,
       1968)</bibl>
  </sourceDesc>
 </fileDesc>
```

スクロールして「一括表示(Show all)のリンクをクリックする

ガイドラインの使用方法（5）

「teiHeader」を一括表示させた画面の一部

2.1.1 The TEI Header and Its Components

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>
<!-- title of the resource -->
   </title>
  </titleStmt>
  <publicationStmt>
   <p>
<!-- Information about distribution of the resource -->
   </p>
  </publicationStmt>
  <sourceDesc>
   <p>
<!-- Information about source from which the resource derives -->
   </p>
  </sourceDesc>
 </fileDesc>
</teiHeader>
```

目録 ⚓

<teiHeader> (zh-TW)

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>Taisho Tripitaka, Electronic version, No. 251 般若波羅蜜多心經
</title>
   <respStmt>
    <resp>底本來源</resp>
    <name>大藏出版株式會社</name>
   </respStmt>
   <respStmt>
    <resp>電子版本製作</resp>
    <name>中華電子佛典協會</name>
   </respStmt>
  </titleStmt>
  <publicationStmt>
   <distributor>中華電子佛典協會</distributor>
   <address>
    <addrLine>11246 台北市北投區光明路276號4樓</addrLine>
   </address>
   <availability>
    <p>本資料使用限定於非營利性用途，並需附上本標頭資訊。</p>
   </availability>
   <date>1998年12月</date>
  </publicationStmt>
  <sourceDesc>
   <bibl>Taisho Tripitaka Vol. 08, Nr. 251 般若波羅蜜多心經</bibl>
  </sourceDesc>
 </fileDesc>
 <encodingDesc>
  <projectDesc>
   <p>預備格式轉換</p>
  </projectDesc>
 </encodingDesc>
 <revisionDesc>
  <list>
   <item>
    <date>19990721/18:35:54</date>CW以CBXML.BAT (99/6/30)轉換成
XML</item>
  </list>
 </revisionDesc>
</teiHeader>
```

一括表示 ⚓

テキストの構造を考える

（目録）
題名　　私の論文
著者　　山田太郎
出版者　阪大出版
出版年　2024

（本文テキスト）
私の論文
山田太郎

序論
これは私の論文の序論です。
…

1. テキストの大まかな構造を把握する

- 書誌情報 --- teiHeaderタグ
- 本文テキスト --- textタグ
  - 章
  - 段落
  - 文の区切り
  - 注釈
  - 引用 など

必要最低限のタグから始めてみる

基本的なタグを使って、テキストの構造を
大まかに表現してみよう

（目録）
題名　　私の論文
著者　　山田太郎
出版者　阪大出版
出版年　2024

（本文テキスト）
私の論文
山田太郎

序論
これは私の論文の序論です。
…

簡単化したマークアップの例

```
<TEI>
  <teiHeader>
      <title>私の研究論文</title>
      <author>山田太郎</author>
      <publisher>阪大出版</publisher>
      <date>2024</date>
  </teiHeader>
  <text>
    <body>
      序論
      これは私の研究論文の序論です。
    </body>
  </text>
</TEI>
```

マークアップの結果例

```
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>私の研究論文</title>
        <author>山田太郎</author>
      </titleStmt>
      <publicationStmt>
        <distributor>阪大出版</distributor>
        <date>2024</date>
      </publicationStmt>
      <sourceDesc>
        <p>この動画のために作成した例</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
```

（<text>に続く）

```
  <text>
    <body>
      <div>
        <head>序論</head>
        <p>これは私の研究論文の序論です。</p>
      </div>
    </body>
  </text>
</TEI>
```

クイックリファレンス

日本語のクイックリファレンスには
こちらのサイトをおすすめします

上位

textcrit: lem rdg
textstructure: back body div front

下位

analysis: interp interpGrp span spanGrp
certainty: certainty precision respons
cmc: post
core: bibl biblStruct cb cit desc divGen ellips
milestone note noteGrp p pb q quote sa
dictionaries: entry entryFree superEntry
drama: camera caption castList move sound
figures: figure notatedMusic table

太字は**モジュール**という、特定の目的に合わせたタグや構造をまとめたパーツのこと

「TEIガイドライン定義モジュール・要素・属性一覧」
https://tei-lab.github.io/element_list/

モジュール一覧

TEIガイドラインのすべての要素は，その用途や使用分野に応じ，以下のモジュールのいずれかに分類されている．モジュール名をクリックすると，要素一覧中の当該モジュールグループの先頭に移動する．

モジュール名	日本語訳モジュール名	ガイドラインの対応する章
tei	全TEIモジュールで使用可能なデータ型，クラス，マクロ．	1. The TEI Infrastructure
header	ヘダーモジュール	2. The TEI Header
core	コアモジュール	3. Elements Available in All TEI Documents
textstructure	テキスト構造モジュール	4. Default Text Structure
gaiji	外字モジュール	5. Characters, Glyphs, and Writing Modes
verse	韻文モジュール	6. Verse
drama	舞台芸術モジュール	7. Performance Texts
spoken	発話モジュール	8. Transcriptions of Speech

村並秀昭「TEI ガイドライン実験室」より抜粋

日本語TEI関連書籍・コミュニティ

1. 一般財団法人人文情報学研究所ほか. 人文学のためのテキストデータ構築入門: TEIガイドラインに準拠した取り組みにむけて. 東京, 文学通信, 2022, ISBN978-4-909658-84-5.
2. TEI研究会　https://tei.dhii.jp/home
3. 青空文庫テキストのTEI自動化プロジェクト
   https://github.com/TEI-EAJ/auto_aozora_tei

おわりに

TEIタグは、人文学系の資料が多種多様であるため、**その資料を使う人やコミュニティがタグ付けを考えるべき**という考え方で運用されている。

そのため取り扱うテキストの種類に近いTEIサンプルを探したりハンズオンでワークするのが手っ取り早くておすすめ👍