

民事第一審判決書の XML データ化

前田 郁勝¹ 外山 勝彦² 小川 泰弘³

¹元名古屋高裁判事 ²名古屋大学大学院情報学研究科 ³名古屋市立大学データサイエンス学部

maeda.ikumasa@gmail.com toyama@is.nagoya-u.ac.jp ogawa@ds.nagoya-cu.ac.jp

概要

民事判決書は現時点で、そのごく一部が最高裁判所や各省庁のウェブサイトから PDF 形式や HTML 形式のデータで公開されているに留まっている。それらのデータ形式は、判決書の検索や解析を行うために適したものとはいえない。また、法務省は民事判決書の全件データベース化に向けて検討を進めているが、データの再利用性や機械可読性の観点から、どのようなデータ形式が望ましいかを検討する必要がある。そこで、民事判決書の文書構造を踏まえて、その XML データ化の手法を示すとともに、最高裁判所ウェブサイトから取得した民事第一審判決書 (PDF 形式) の XML データ化を行った。

1 はじめに

我が国における民事判決書は、現時点で、最高裁判所ウェブサイトが地裁、高裁、最高裁の判決書データを PDF 形式で公開しているほか、一部省庁ⁱ がそれぞれ所管分野の判決書データを PDF 形式や HTML 形式で公開している。また、民間業者により判例検索サービスが有償で提供されている。

判決書を適切に検索したり、その高度な解析を実行したりするためには、公開される判決書データは機械処理が容易なデータ形式が望ましい。

この点について、「民事司法制度改革の推進について」(令和 2 年 3 月 10 日民事司法制度改革推進に関する関係府省庁連絡会議)ⁱⁱ において、民事判決情報を広く国民に提供することの意義として、国民に対する司法の透明性向上、国民に対する行動規範・紛争解決指針の提示、紛争解決をサポートする AI の研究・開発等の意義が指摘された。また、民事訴訟手続の改正 (令和 4 年法律第 48 号) により、判決書が電磁的記録をもって作成されることが定められた (改正後の民訴法 160 条 1 項, 同 252 条 1 項)。法

務省は「民事判決情報データベース化検討会」ⁱⁱⁱ を設置し、民事判決書の全件データベース化や、その際の仮名化処理や情報管理の在り方等を検討している。

民事判決書の全件公開では、その数は年間約 20 万件になると想定される。その際に、再利用や機械可読が容易な形式で判決書データが提供されれば、その解析や利用が進み、司法機能の効率化や司法サービスの充実に貢献することが期待される。

本稿では、データの再利用性や機械可読性の観点から、民事判決書の文書構造を踏まえた XML データ化の手法を示すとともに、最高裁判所ウェブサイト中の「裁判例」(以下「最高裁サイト」という。)^{iv} から取得した民事第一審判決書 (PDF 形式) の XML データ化を行ったことを報告する。

なお、裁判所が扱う事件には、民事事件、行政事件、刑事事件、家事事件等の種類があり、それぞれの事件における裁判には判決、決定、命令の形式があるが、本研究においては、全件公開に向けて準備が進められている私人間の紛争を扱う民事判決を対象とする。また、第一審、第二審、第三審の判決のうち、件数が最も多い第一審判決を対象とする。

以下、2 節で民事判決書の文書構造について、3 節で XML 文書スキーマの設計概要についてそれぞれ述べる。4 節では本稿のまとめと課題を述べる。

2 民事判決書の文書構造

本節では、民事判決書の文書構造の概要を説明した後、その各部分構造について詳細に述べる。

2.1 判決書の文書構造の概要

民事判決書の様式には、大きく分けて、平成 2 年頃までの「在来様式」と、その後提唱された「新様式」の 2 つがある [1] [2]。

新様式判決書 (図 1) は、冒頭から順に、事件の特

ⁱ 国税庁：税務訴訟資料，総務省：情報公開・個人情報保護関係資料，法務省：訟務重要判例集，厚生労働省：労働基準関係判例，労働委員会関係の命令・判決書

ⁱⁱ <https://www.cas.go.jp/jp/seisaku/minjikaikaku/>

ⁱⁱⁱ https://www.moj.go.jp/shingi1/shingi09900001_00004.html

^{iv} https://www.courts.go.jp/app/hanrei_jp/search1

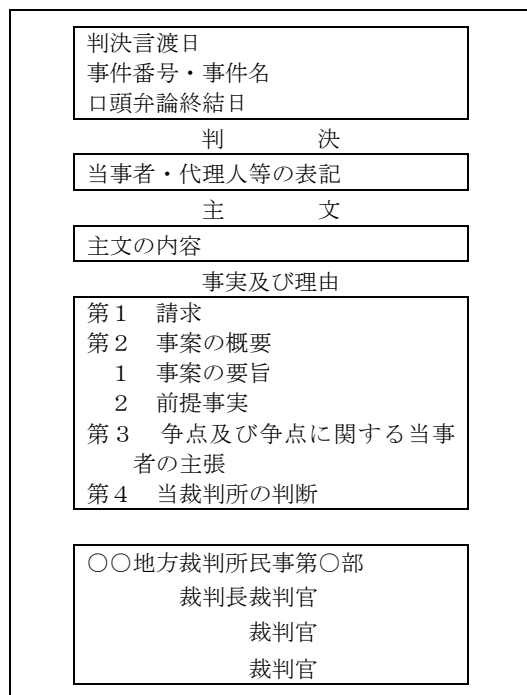


図 1 新様式判決書の文書構造 (概要)

定に関する事項 (日付や事件番号・事件名), 「判決」などの標題, 事件に関わった当事者・代理人等が記載された後, 判決の結論を示す「主文」, 当事者の請求, 主張, 裁判所の判断理由が記載される「事実及び理由」が続き, 末尾に裁判所・裁判官名が記載される。「主文」と「事実及び理由」の冒頭にはそれぞれの標題が付される。ここで, 標題「事実及び理由」は, 民訴法 253 条に定められた判決書の記載事項の「事実」及び「理由」を意味するものであり, 「事実」の記載においては, 当事者の請求と主文が正当であることを示すために必要な主張を摘示することが求められる。

一方, 在来様式の判決書は, 新様式判決書の「事実及び理由」の部分で「事実」と「理由」に分けて記載していた (図 2)。現在では, 新様式による判決書が定着し, 新様式の方が在来様式よりも圧倒的に多い。そこで, 本稿では新様式判決書を対象とする。

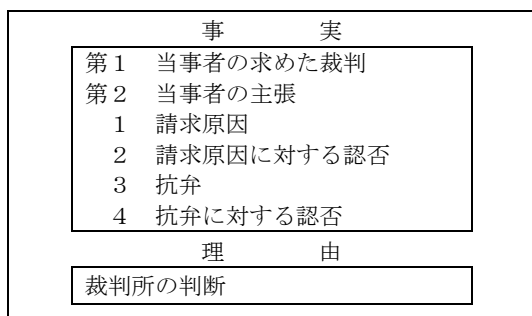


図 2 在来様式判決書の文書構造 (一部)

2.3 新様式判決書の部分文書構造

新様式判決書の部分構造について補足する。

1 冒頭部分及び末尾部分

冒頭部分の判決言渡日, 事件番号・事件名, 口頭弁論終結日, 当事者の表記, 「主文」の標題とその内容, 末尾部分の裁判所名・裁判官名の記載は, その標題などから記載内容が明瞭であり, また, 判決書によって記載形式に大きな差異はない。

なお, 判決言渡日は, 裁判所書記官が記載すべき事項であるが (民訴規則 158 条), 判決書中に一体として記載される場合が多いことから, 判決書データの一部として扱うことにする。

2 「事実及び理由」部分

この部分は, 概ね次の順に記載される。

- (1) 請求: 原告が被告に求めた請求の内容が記載される。
- (2) 事案の概要: 事案の要旨と前提事実等が記載される。
 - ・事案の要旨: 紛争の概要や原告の権利主張の内容が簡潔に記載される。標題として「事案の要旨」を記載する場合もあるが, そうでない場合も多い。
 - ・前提事実: 当事者間に争いのない事実や証拠により明らかに認められる事実を記載する。「前提事実」, 「争いのない事実」などの標題が記載され, 標題は判決書によって異なることがある。
- (3) 争点及び争点に関する当事者の主張: 事件によって, 争点が 1 つの場合もあれば, 複数の場合もある。後者の場合には, 争点ごとに当事者の主張が記載される。また, 「争点」と「争点に関する当事者の主張」のそれぞれに標題を設定する場合もある。当事者の主張部分の標題として, 「(1) 原告の主張」などと記載する場合もあれば, 「(原告の主張)」などと括弧書きにする場合もある。
- (4) 当裁判所の判断: 争点に対する裁判所の判断を記載する部分である。全体の事実経過をまとめて事実認定して物語風に記載したうえで, 各争点に対する判断を記載する場合もあれば, 争点ごとに, その判断に必要な限度で事実認定を記載する場合もある。

3 XML 文書スキーマの設計

本節では, 民事判決書の文書構造に基づき, その XML 文書スキーマの設計について述べる。

3.1 基本方針

本研究では、文書指向 (document-centric) の文書構造化を採用する。すなわち、民事判決書という文書に対して、基本的には、文や段落、またはそれらのまとまりなど、文書の表層構造に注目して文書構造化を行う。一方、各文や各段落などが判決の結論に至る議論の中で果たす種々の意味的役割を要素として直接的に定義する文書構造化の方法も考えられる。そのような役割の付与は判決書の機械処理のために有用である。しかし、そのような要素の同定のためには、テキストの意味解析が必要になる場合が考えられる。また、その場合でも文書の表層構造が前提になる。意味的な情報の抽出は、表層から直接的に得られるものに留める。

また、前節で述べたように、判決書は文書構造の自由度が高い。法令の文書構造化では、その文書構造が厳格に定まっていることを利用して、デジタル庁・法令標準 XML スキーマ^v が設計されたが、判決書の文書構造化では、この自由度を考慮した文書スキーマの設計が必要になる。

民事判決書の基本的な文書構造は、前節で述べたとおりである。そのうち冒頭部分と末尾部分は、この判決に関するメタ情報の記載であり、両者の間が判決の内容に関する判決書の本体部分である。そこで、判決書全体に対して要素 **Decision** を定義し、判

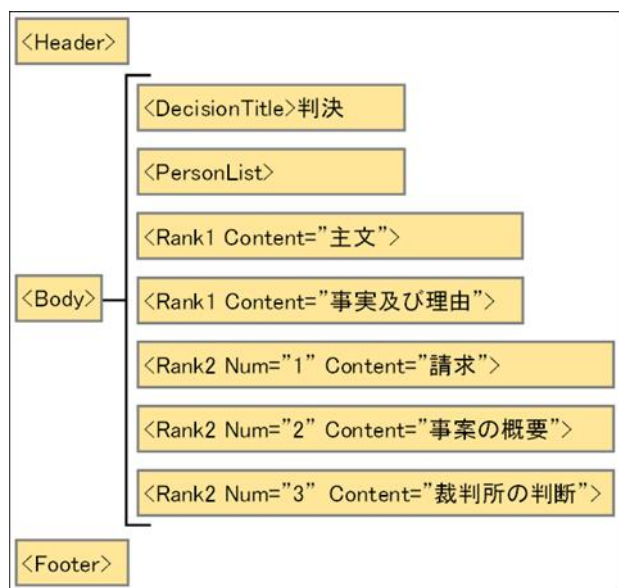


図 3 民事判決書の XML スキーマ (概要)

決書の冒頭部分、本体部分、末尾部分に相当する子要素として、Header, Body, Footer の 3 つを定義した (図 3)。以下、それらを順に説明する。

3.2 冒頭部分 Header

判決書の冒頭部分に記載された①判決言渡日、②事件番号・事件名、③口頭弁論終結日の部分は、それぞれ要素 **Header** の子要素とした。また、それらの具体的内容は、判決書を特定するために必要であり、検索時に利用する可能性が高いので、それぞれ、それらの子要素の属性として記述することにした。図 4 に例を示す。

なお、最高裁サイトから取得できる判決書データの中には、上記①、②、③の記載がないものがあるが、それらの情報が最高裁サイトから別途取得できる場合は、それをを用いて属性値だけを補い、開始タグと終了タグの間は空文字とした。

3.3 本体部分 Body

判決書の本体部分は、「判決」などの標題、当事者・代理人等、主文、事実及び理由の各部分から構成される。以下、各部分の構造化について述べる。

(1) 「判決」などの標題

要素 **Body** の子要素として、**DecisionTitle** を定義した。なお、この標題のテキストは、「中間判決」「追加判決」「手形判決」「少額訴訟判決」など、判決の種類に応じたものとなる。

(2) 当事者・代理人等

要素 **Body** の子要素として、**PersonTitle** と **PersonName** を定義した。前者は「原告」「被告」「参加人」「訴訟代理人弁護士」など、当事者・代理人等のタイトルに対して、また、後者はその氏名や会社名を表す。

①	<code><Render Era="Heisei" Year="29" Month="6" Day="1">平成29年6月1日判決言渡し 同日原本交付 裁判所書記官</Render></code>
②	<code><CaseID Era="Heisei" Year="28" Code="ワ" Number="9999" Name="不正競争行為差止等請求事件">平成28年(ワ)第9999号 不正競争行為差止等請求事件</CaseID></code>
③	<code><ArgumentEnd Era="Heisei" Year="29" Month="4" Day="1">口頭弁論終結日 平成29年4月1日</ArgumentEnd></code>

図 4 要素 Header の子要素 (例)

^v https://elaws.e-gov.go.jp/file/XMLSchemaForJapaneseLaw_v3.pdf

なお、「亡山田太郎訴訟承継人 山田花子」のように、氏名「山田花子」の直前や前行に記載される記述「亡山田太郎訴訟承継人」に対して、子要素 `PersonPreNote` を定義した。一方、氏名の次行に記載される「(以下「原告〇〇」という。)」等の記述に対して、子要素 `PersonPostNote` を定義した。

(3) 主文, 事実及び理由

この部分は、階層的な構造により記述される。判決書では公用文作成の要領[3]に従い、文書の部分構造の階層レベルや同レベルでの順序を表すための付番ポリシーとして、「第1」「1」「(1)」「ア」「(ア)」「a」「(a)」という付番表現がこの順に用いられることが多い。しかし、そのポリシーに従っていない判決書は少なくない。また、前節で述べたとおり、事案の要旨、当事者の主張などでは、その付番ポリシーは一定してない。

法令標準 XML スキーマにおいては、法令の階層構造(編, 章, 節, 款, 目, 条, 項, 号, 号の細分など)を表すための付番ポリシーは厳格であるため、付番表現(「第一条」, 「2」など)をもとに要素の階層関係を定義している。しかし、民事判決書においてはそのような厳格性は見られないので、階層や順序を表すための付番表現をそのまま用いて構造化することは好ましくない。

そこで、付番表現によって各部分構造の範囲を特定したうえで、その階層レベルの浅い順に `Rank1`, `Rank2`, `Rank3`, … という子要素を機械的に定義することにした。また、同時に、各部分構造に出現する重要な標題(主文, 事実及び理由, 請求, 事案の概要, 事案の要旨, 争点, 当事者の主張, 裁判所の判断等)を属性 `Content` の値として付与し、その部分構造が判決書の中で果たす役割を明らかにすることにした。各 `Rank` に付番表現や標題の記載がある場合は、それぞれ子要素として定義した(`Rank2` に対しては, `Rank2NumTitle`, `Rank2ContentTitle`)。図 5 は、「事実及び理由」に続いて「第1 請求」と記載されている部分を構造化した例である。

```
<Rank1 Content="事実及び理由">
  <Rank1ContentTitle>事実及び理由</Rank1ContentTitle>
  <Rank2 Num="1" Content="請求">
    <Rank2NumTitle>第1</Rank2NumTitle>
    <Rank2ContentTitle>請求</Rank2ContentTitle>
  </Rank2>
</Rank1>
```

図 5 本体部分 Body の階層構造 (例)

3.4 末尾部分 Footer

判決書の末尾部分には、裁判所や裁判官に関する情報が記載される。要素 `Footer` の子要素は次のように定義した。

(1) 裁判所

裁判所に関する記載に対して子要素 `Court` を定義し、その属性として、裁判所の種類 `CourtType`, 裁判所の名前 `CourtName`, 裁判所内部の部番号 `CourtDivision` または支部名 `CourtBranch` を定義した。

(2) 裁判官名

裁判官名に関する情報を示す子要素 `Judge` を定義した。さらにその子要素として、裁判官の肩書と氏名を表す `JudgeTitle` と `JudgeName` を定義した。

4 XML データ化作業

現在、最高裁サイトから取得した民事第一審判決書データ(PDF形式)に対して、テキストデータにいったん変換したうえで、前節で述べた文書スキーマに則って、XML形式に変換する作業を進めている。すでに作業を終えたXMLデータはスキーマとともに公開予定であり、今後も作業の進展のつど、データを追加していく。

なお、最高裁サイトから取得したPDF形式データでは、外字は画像データになっている。3.3節で述べた付番表現が外字である場合は、テキストデータ(「(1)」「(ア)」「(a)」など)に変換した。また、氏名の一部が外字であって、それがUnicodeに存在しない場合は、近似する文字に変換するとともに、当該文字に<Variant>タグで囲む措置を施した。

5 おわりに

本稿では、再利用性と機械可読性のある判決書データの提供・公開を目的として、民事第一審判決書のXML文書スキーマの設計について述べた。現在、要素 63 種類、属性 15 種類を定義している。

なお、在来様式判決書についても、付与する属性を在来様式に合わせることにより、同様にXMLデータ化できる。また、高裁判決、最高裁判決の判決書や民事判決以外の判決書についても、同様にXMLデータ化できると考えられる。それらについては引き続き検討する。

謝辞

本研究の一部は JSPS 科研費 JP21H03772, JP22H03901 の助成を受けたものです.

参考文献

- [1] 司法修習所 (編): 民事判決起案の手引き (10 訂 (補訂版), 法曹界 (2020) .
- [2] 最高裁判所事務総局 (編): 民事判決書の新しい様式について, 法曹界 (1990) .
- [3] 文化庁 (編): 新訂 公用文の書き表し方の基準 (資料集), 第一法規 (2011) .

付録

```

<?xml version="1.0" encoding="utf-8"?>
<Decision>
  <Header>
    <Render Era="Heisei" Year="29" Month="6" Day="1">平成29年6月1日判決言渡し 同日原本交付 裁判所書記官</Render>
    <CaseID Era="Heisei" Year="28" Code="7" Number="9999" Name="不正競争行為差止請求事件">平成28年(ワ)第9999号 不正競争行為差
止請求事件</CaseID>
    <ArgumentEnd Era="Heisei" Year="29" Month="4" Day="1">口頭弁論終結日 平成29年4月1日</ArgumentEnd>
  </Header>
  <Body>
    <DecisionTitle>判決</DecisionTitle>
    <PersonList>◇当事者・代理人リストの記載は省略◇</PersonList>
    <Rank1 Content="主文">
      <Rank1ContentTitle>主文</Rank1ContentTitle>
      <Rank2 Num="1">
        <Rank2NumTitle>1</Rank2NumTitle>
        <Paragraph Num="1">
          <Sentence Num="1">原告の請求を棄却する。</Sentence>
        </Paragraph>
      </Rank2>
      ◇主文2項(訴訟費用負担の裁判)の記載は省略◇
    </Rank1>
    <Rank1 Content="事実及び理由">
      <Rank1ContentTitle>事実及び理由</Rank1ContentTitle>
      <Rank2 Num="1" Content="請求">
        <Rank2NumTitle>第1</Rank2NumTitle>
        <Rank2ContentTitle>請求</Rank2ContentTitle>
        ◇請求の記載は省略◇
      </Rank2>
      <Rank2 Num="2" Content="事案の概要">
        <Rank2NumTitle>第2</Rank2NumTitle>
        <Rank2ContentTitle>事案の概要等</Rank2ContentTitle>
        <Rank3 Num="1" Content="事案要旨">
          <Rank3NumTitle>1</Rank3NumTitle>
          <Rank3ContentTitle>事案の要旨</Rank3ContentTitle>
          <Paragraph Num="1">
            <Sentence Num="1">本件は、原告が、被告において原告と同一又は類似の営業表示を使用しているとして、被告に対し、不正競争防止法2条1
項1号、3条1項に基づき、被告が「△△△△」を含む表示を使用することの差止めを請求した事案である。</Sentence>
          </Paragraph>
        </Rank3>
        <Rank3 Num="2" Content="前提事実">
          <Rank3NumTitle>2</Rank3NumTitle>
          <Rank3ContentTitle>前提事実</Rank3ContentTitle>
          ◇前提事実の記載は省略◇
        </Rank3>
        <Rank3 Num="3" Content="争点">
          <Rank3NumTitle>3</Rank3NumTitle>
          <Rank3ContentTitle>争点</Rank3ContentTitle>
          ◇争点の記載は省略◇
        </Rank3>
        <Rank3 Num="4">
          <Rank3NumTitle>4</Rank3NumTitle>
          <Rank3ContentTitle>争点についての当事者の主張</Rank3ContentTitle>
          ◇当事者の主張の記載は省略◇
        </Rank3>
      </Rank2>
    <Rank2 Num="3" Content="裁判所の判断">
      <Rank2NumTitle>第3</Rank2NumTitle>
      <Rank2ContentTitle>当裁判所の判断</Rank2ContentTitle>
      ◇裁判所の判断の記載は省略◇
    </Rank2>
  </Body>
  <Footer>
    <Court CourtType="District" CourtName="大阪" CourtDivision="第21民事部">大阪地方裁判所第21民事部</Court>
    <Judge>
      <JudgeTitle>裁判長裁判官</JudgeTitle>
      <JudgeName>〇〇〇〇</JudgeName>
    </Judge>
    ◇他の裁判官2名の肩書、氏名の記載は省略◇
  </Footer>
</Decision>

```

図 A 判決書 XML データ (例)