

## 特許事務処理システムのXML化

特許庁情報システム課  
安久 司郎

### 1. はじめに

特許庁では1990年から特許電子出願を基礎としたペーパーレスシステムを稼働し、出願手続き、庁内事務処理および審査、特許公報の発行、電子図書館の構築等の一連の業務の流れを電子データによる処理で行える環境を構築し、実際にも業務の効率化を進める中核となってきた。このペーパーレスシステムは、1985年当時のシステム設計段階で考えられる技術の中で最新かつ世界標準となる可能性の最も高いものを採用し、将来に渡ってシステムが維持できることを目標にしたものであった。例えば、データ通信プロトコルはOSIを採用し、電子文書の仕様は当時のCCITT規定のミクストモードデータ文書交換標準仕様をベースに構築していた。しかしながら、今やこれらの技術のデファクトスタンダードはWeb技術、HTTPをベースとしたインターネット環境であり、また、電子文書の標準もXMLが大きく注目され、Webとの連携の上では必須のものとなりつつある。そこで、これらの新たな技術に対応するために、まずは2003年からペーパーレスシステムで扱う電子文書仕様をXML対応とし、出願から特許公報までの一連の処理を、XMLをベースとした形態で行えるようにすることとした。

### 2. 特許XMLデータの世界標準の策定

XMLを採用するにあたり、最初に検討しなくてはならない点は、その標準化ということである。特許の世界の場合、扱う文書データは、出願人、発明者等の氏名、住所に代表される書誌的データと特許の技術内容を説明する明細書データに大別されることになる。書誌的データについては、特許法に規定された非常に複雑な手続き処理を行う必要があるため、特許法および規則で規定された事項に関する非常に多くのボキャブラリが必要となる。また、明細書の場合、通常のテキストデータの他に、図面、表、数式、化学式、遺伝子配列等の多岐にわたる種類のデータを扱う必要がある。

これらの内容について、標準化を行い使用するXMLを規定することになったが、今や特許データの利用は日本国内で閉じたものでは全くなく、全世界的に通用するものでなくてはならず、標準化にあたっては、世界標準として策定し、それを採用する必要がある。そのため、世界知的所有権機関(WIPO)を中心として、この標準が取りまとめられることとなり、日本特許庁もその主要メンバーとして標準策定に参加した。

この国際標準策定にあたって、基本的に対象とするモデルとしては、PCT(特許協力条約に基づく国際出願)の様式が基礎とされた。その理由としては、現在、各国はそれぞれ独自の特許法を持っているため、使用される書類の様式も各国特許庁ごとに大きく異なり、それら全てを網羅したボキャブラリ、スキーマ等を策定するのは容易なことではないので、現状では、唯一、全世界で統一様式として利用されているPCT国際出願の様式により、基本のボキャブラリ、スキーマを策定した。そのため、本標準の中には、この標準を各国で使用する場合のこの標準の国内仕様への拡張方法に関するルールも規定することとした。

この標準は、WIPOにより特許協力条約実施規則付属書F<sup>(1)</sup>として2001年12月に発効し、現在は、PCT国際出願を電子的に行う場合については、各国特許庁が守らなくてはならない基準書となっている。この標準の中でXMLに関して規定されたものの概要は以下のとおりである。

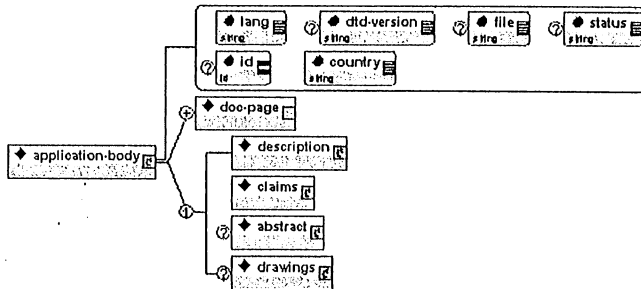
(1) 文書DTD

中核をなす DTD は以下のものがある。(最新のバージョンでは以下のものを含め、約 20 の DTD が規定されている)

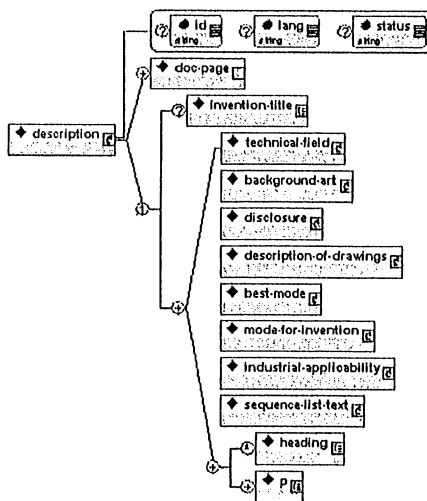
- package-data.dtd : 文書パッケージエンベロープ用 DTD
- request.dtd : 願書データ DTD
- fee-sheet.dtd : 料金表 DTD
- application-body.dtd : 出願明細書等 DTD

application-body は明細書データであるため、文字飾りを含むテキスト、表、数式、図面等に対応しており、作成にあたっては以下の点が考慮されている。

1. 使える部分については、XHTML で定義されているコンポーネントを利用
2. PCT 条約で規定されている項目は独自タグとして規定
3. 数式は W3C 規定の mathml2 をモジュールとして組み込む (オプション)
4. 表は OASIS の Open XML Table Model を組み込む (オプション)
5. 外部参照実体として、イメージファイル (形式は TIFF、JFIF、PDF) を参照可能



application-body.dtd の全体概要



application-body 内の明細書部分の概要 (例えば、p はパラグラフ用のタグ)

(2) 文字コード

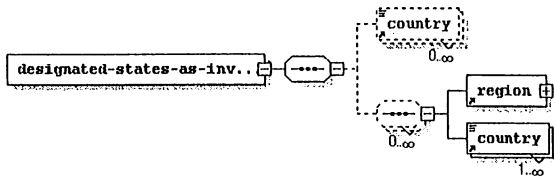
全世界での利用を考慮し、文字コードセットは Unicode3.0、エンコードは UTF-8 を標準とした。ただし、各国ごとに他のエンコードの利用も許されている。(例：日本の場合は Shift-JIS)

(3) スタイルシート

PCT 国際出願用に、標準のスタイルシートが規定されている。PCT 規則の関係で、明確なページレイアウトが要求されているため、XSL-FO によりそれぞれの文書スタイルシートが規定されている。

この PCT 国際出願の標準仕様内で定義された DTD に基づいて、標準スキーマコンポーネントセットの定義も別途行っている。このコンポーネントは、XML Schema を利用して、個々のタグを分解し、コンポーネントとして定義し直したものである。新たな文書 Schema を作成する場合を、このコンポーネントを組み合わせることで作成することができる。

以下は、"designated-states-as-inventor" というコンポーネントの例で、子供のコンポーネントとして "country" と "region" を別の XML Schema コンポーネントとして取り込んでいる。



```
<xsd:schema xmlns:xsd = "http://www.w3.org/2001/XMLSchema">
  <xsd:include schemaLocation = "country.xsd"/>
  <xsd:include schemaLocation = "region.xsd"/>
  <xsd:element name = "designated-states-as-inventor">
    <xsd:annotation>
      <xsd:documentation>
        PCT Rule 4.6(c) :
        "The request may, for different designated states, indicate different persons as inventors where, in
        this respect, the requirements of the national laws of the designated States are not the same. In such
        a case, the request shall contain a separate statement for each designated State or group of States in
        which a particular person, or the same person, is to be considered the inventor, or in which particular
        persons, or the same persons, are to be considered the inventors."
      </xsd:documentation>
    </xsd:annotation>
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref = "country" minOccurs = "0" maxOccurs = "unbounded"/>
        <xsd:sequence minOccurs = "0" maxOccurs = "unbounded">
          <xsd:element ref = "region"/>
          <xsd:element ref = "country" maxOccurs = "unbounded"/>
        </xsd:sequence>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
</xsd:schema>
```

```

</xsd:sequence>
</xsd:sequence>
</xsd:complexType>
</xsd:element>
</xsd:schema>

```

現在のところ、XML Schema については、この標準コンポーネントセットの定義にのみ使用しており、文書スキーマはあくまでもDTDを基本としている。新たな文書DTDを作成する場合は、まず、このXML Schema コンポーネントを組み合わせることで1つの文書用のXML Schema を作成し、最後にDTDに変換する手法を用いている。

### 3. 日本国内特許データへの適用

日本国特許庁においては、この標準コンポーネントから国内特許出願用のDTDの作成を行っている。この場合、国内出願で規定されている規則に対応した独自のタグが必要となるため、標準化の作業の中で作成されたルールに基づき、ネームスペースを利用して独自のタグを定義している。ただし、DTDはネームスペースを基本的にはサポートしていないので、現状では、DTD内に `jp:` というプレフィックスを最初から埋め込むことで実現している。以下は、サンプルDTDの一部である。

```

<!ELEMENT jp:pat-amnd (jp:amendment-a51 | jp:amendment-a523 | jp:amendment-a524 |
    jp:amendment-a525 | jp:amendment-a526 | jp:amendment-a527 | >
<!ATTLIST jp:pat-amnd
    lang CDATA #REQUIRED
    dtd-version CDATA #IMPLIED
    file CDATA #IMPLIED
    status CDATA #IMPLIED
    xmlns:jp CDATA #FIXED "http://www.jpo.go.jp" >
<!--information regarding Agents or common representatives JPO expanded-->
<!ELEMENT jp:agents (customer-number | jp:agent+ ) >
<!ATTLIST jp:agents
    jp:error-code CDATA #IMPLIED >

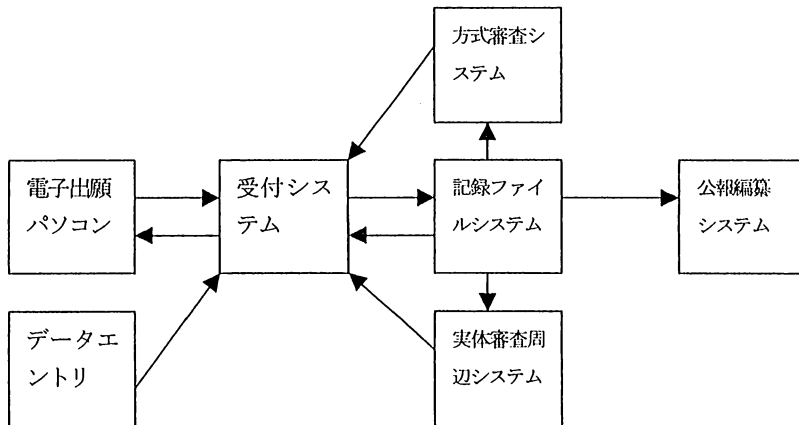
```

なお、日本特許庁では、`mathML2` については現状では一般的なツールがまだ十分でないとの判断から、採用を見合わせており、複雑な数式は当面はイメージデータを利用することとなっている。また、スタイルシートについては、国内出願の場合はXSLTによるHTML変換を利用している。(PCT国際出願についてはXSL-FOを利用しているが、出願件数の多い国内出願では現在のところ利用していない。)

### 4. 特許庁内事務処理システムへの実装

特許庁の事務処理システムの大まかな流れは下図の通りであり、すでに全ての処理は電子データで行わ

れている。電子出願の受付、もしくは、書面による出願を XML データ化したものを受け付けシステムが受領し、各種処理を行った後に、電子公報として同様に XML で発行されるまでの流れを対象としている。これらのシステムを全て、これまで述べてきた標準に準拠した XML 対応のシステムに変更、もしくは新規システムとして開発を行った。



- 電子出願パソコン (パソコン出願ソフト)

従来は、CCITT 規定のミクストモードデータ文書交換標準仕様をベースとした独自のフォーマットを用いて電子出願を行っていたが、2003年7月から、その基本フォーマットを XML に変更し、業務運用を開始した。出願人は、出願データを作成するにあたり、はじめに HTML 形式で願書や明細書を作成し、パソコン出願ソフトが持っている HTML→XML 変換ツールで日本特許庁仕様の XML にそれぞれの文書データを変換する。必要な XML を生成した後、それらの XML と添付されるイメージファイル等の外部ファイルをパッケージ情報管理 XML と共に ZIP により 1 ファイルとし、ISDN 回線を通して、パソコンから特許庁の受付サーバに対して送信、電子出願を行う。

また、特許庁で作成された各種通知類も XML データとして作成され、同様の形式により電子出願パソコン側で受領することが可能である。

- データエントリ (書面出願対応)

特許出願は、電子出願によるもののみではなく書面によるものも受領する必要があるため、受領した書面出願の内容を電子出願と同様な形式の XML データに変換して、その後の処理を行っている。書面による出願は、特許庁での受領後 OCR によりコードデータに、図面はイメージデータに変換し、規定されているタグを付与して標準の XML を生成する。このデータを受付システムに投入して、電子出願されたものと同様の処理を行う。

- 受付システム

ISDN を通じて受領した XML ファイルを含む ZIP ファイルは受付システムで解凍され、DTD によるパースチェック、その他のデータ形式チェックを行い、後続システムに引き渡される。また、受付システムは、特許庁内の方式審査システム、実体審査周辺システムで作成される各種の通知、起案類の XML データを電子出願パソコン側に渡す機能も実装しており、オンラインによる発送、書類閲覧も XML により行うことができる。

- 記録ファイルシステム

XML データとして受領した出願データは、全て記録ファイルシステムで管理される。このシステムで記録、保管される XML データが、その特許出願の原本データとなるため、それぞれの書類ごとに XML データを管理サーバ内のデータベースにそのまま蓄積し、管理を行っている。また、XML データ内に存在する各種の主要データ項目（出願人名、住所、代理人情報等）は XML データから抽出され、別途データベースが構築される。これらのデータベース内の情報を、後続の処理システムが利用することとなる。なお、これらの XML データによる文書を照会する場合、照会用の PC において、対象となるスタイルシートを適用して、規定されたレイアウトで照会することができる。

- 方式審査システム、審査周辺システム

方式システムは、XML による出願データが法律、規則に適合しているかのチェックを行うための支援システムである。XML データ受領直後に、受付システムでパースチェック等の基本的なチェックは行われているが、この方式審査システムでは、タグ内に記載された内容同士の対比チェックや、他の XML 文書データ内の項目との対比による妥当性チェックもシステム上可能な範囲で自動的に行っている。方式上の違反がある場合は、その指摘内容を記載した通知類を担当官が XML で作成し、出願人が受領できるようになっている。また、実態審査周辺システムでは、審査官が特許審査に関する各種の通知文書を同様に XML で作成し、出願人に受付システムを通して送付している。

- 公報編纂システム

特許出願されたものは、基本的には全て公開公報という形で公開され、さらに、特許となった場合には特許公報という形態で、その特許された内容が発行される。これまで、簡易な SGML 形式で、これらの公報は電子的に CD-ROM で発行されてきたが、今回の XML 化に伴って、これらの公報も XML 化し、DVD-ROM による発行を行う予定である。(2004 年 1 月を予定) XML 公報を発行する場合、提出された XML データをそのまま発行するわけではなく、各種の編集が必要となる。そのため、公報用の DTD は出願用に定義されたものとは異なる別の DTD を使用しており、それらは出願用 DTD よりも多くの付加情報を含んでいる。また、特許庁における審査の段階で、出願人はその出願内容の補正を行うことが認められているため、公報編纂時にはそれらの補正された内容を差し替えて編集する必要がある。例えば、ある段落をそのまま補正された新たな内容に入れ替えたり、指定された図面を入れ替えたり、削除するといったものである。これらの編集について、今回の XML 公報編纂システムでは XML の特質を極力生かして、可能な限り自動的に行うことを目指している。

## 5. 今後の展開

ここまで述べたように、国内出願に関しては、基本的には従来からのシステムの基本文書仕様を XML に置き換え、新たなシステムの運用を実施することができた。今後はこれらの機能をさらにレベルアップしてより XML の特徴を生かせるようにシステム化を進めることとなる。また、来年には、そもそもの XML 化の標準化のきっかけとなった国際出願の電子出願開始を予定しており、これが実現した後は、国際出願としての XML 出願データをこの標準に準拠して受領し、世界各国に展開されていくことになり、さらなる発展が期待される。

参考文献・情報

1. WIPO, "Standard for the electronic filing and processing of international applications: Annex F"  
[http://pcteasy.wipo.int/efiling\\_standards/EFPage.htm](http://pcteasy.wipo.int/efiling_standards/EFPage.htm), [http://www.wipo.int/pct/en/texts/pdf/ai\\_01\\_part7.pdf](http://www.wipo.int/pct/en/texts/pdf/ai_01_part7.pdf)
2. Paul Brewin , Shiro Ankyu, "Filing and Processing Patent Data Using XML - A World Standard", XML Europe 2002, Barcelona, Spain.