

(資料)

## 1. XMLとは

XML (Extensible Markup Language 「エックスエムエル」と読む)

### (1) XML 文書の構成

XMLの文書は、次の3つのパートに分かれている。

- ア. XML 宣言
- イ. XML スキーマ
- ウ. 本文 (インスタンス)

XML宣言は、その文書がXML文書であることを示し、そのXML文書がどんな文字コード (UTF-8 やシフト JIS など) で書かれているかの情報が入っている。添付文書情報XMLにおいては、この文字列を必ず文頭に記載する必要がある。

XMLスキーマは、XML文書の文書構造を定義する。例えば、「章の中には一つのタイトルと複数の節がある」「節の中には一つのタイトルと複数の段落がある」という具合である。このXMLスキーマ定義は公開されている物が多くあり、独自の文書構造でないのであれば文書作成者がこの構造を記述する必要はない。又、社内文書等の文書構造が殆ど変更されないような文書は一度作成すれば再利用が可能であるというメリットがある。

本文(インスタンス)は、実際の文書の本文である。「章」等の文書構造や「斜体」等の装飾は、通常の文章に「タグ」という形で付加される。例えば「斜体」のタグを *Italic* として定義した場合、文字の強調は、<Italic>斜体にした文字</Italic>と表現される。<>の間に入っている「Italic」がタグ名である。このタグ名はXMLスキーマ定義で設定する。

以下は、XML 本文(インスタンス)のサンプルである。

```
<book>
  <title>本のタイトル</title>
  <chapter>
    <title>章のタイトル</title>
    章の序文
    <section>
      <title>節のタイトル</title>
      <paragraph>段落 1</paragraph>
      <paragraph>段落 2</paragraph>
    </section>
  </chapter>
</book>
```

## (2) XML の表示 (表現) 方法

XML は出力先を特定していない。文書には、紙出版のための文書や C D - R O M 等に利用する電子出版、インターネットに W W W を経由して公開するネットワーク文書等、様々な出力形態が存在する。又、文書データベースとして活用する方法もある。

例えば、ワードプロセッサ (以下、「ワープロ」という。) は紙出版のみを目的としている。ワープロのデータではどこが章のタイトルで、どこが節のタイトルか、又、どれが強調の意味で使っている斜体文字か、どれが数式の意味で使っている斜体文字か、といった情報は人間が一つ一つ調べない限り、分からない。さらに、「見たまま」を印刷するので、どこで改行するかといったことに神経を使ったり、「タイトルのフォントを変えよう」と思ったときに文書の頭から一つ一つ修正する必要がある。また、ワープロで書いた物を電子出版するのは殆ど作り直しに近くなってしまう。

XML では、文書構造を規定している XML スキーマに従って文章が作成される。例えば、斜体だったら **<Italic>**、上付き文字だったら **<Sup>**、のようになり、表示する時には同じ標準書体でも意味は異なる。「改行の位置」や「タイトルのフォント」といった体裁の情報は、文書で書いた時のものには依存しない。「どこで改行するか」、「タイトルのフォントは何か」といった体裁 (文書の見ばえ) 情報は XML ではなく、XML 文書の「見ばえ」を定義する「X S L T (X S L Transformations) スタイルシート」に任される。

出力先 (普通紙や印刷用版下・C R T など) でどのように文字の装飾や文書の見ばえを表現するか、という体裁情報は「X S L T」で定義する。出力先が紙だったら、紙用の X S L T を、電子ブックだったら電子ブック用の X S L T を、といったように出力先に合わせた X S L T を作成する事により元の XML 文書を全く変更することなく、出力形態に合わせた文書出力を行うことが出来

る。

### (3) XML ファイルの文法エラー

添付文書情報のXMLファイルを提出する場合、提出者（この場合は、製薬会社）は、提出するXMLファイルに文法上エラーがないことを保証しなくてはならない。この保証をする為にXML文法チェッカー（XMLパーサー）を使用して、エラーが発生しないことを保証する必要がある。

XMLパーサーを使用して文法チェックを行うと、様々なエラーが発生することがある。文法エラーとなる主な原因としては、

- ア. タグ（< と >）で囲まれている所に全角文字（全角スペースも含む）や半角カタカナを使用している場合。
- イ. 開始タグと終了タグの1対1対応が崩れている場合。
- ウ. XMLインスタンスに許されていないタグを追加した場合。
- エ. XMLインスタンス（データの記載部分）に許されていない半角文字（「/（半角スラッシュ）」、「&（半角&のみ）」、「<」、「>」等）を使用した場合。

などが考えられる。

XMLパーサーを使用していて文法エラーが出た場合は、エラー原因の検討と、エラーメッセージの内容を解釈して、エラーの原因を特定すること。

注) パーサーはXMLファイルの文法的な誤りを検出するものであり、入力された内容が正しいかどうかを点検するものではない。パーサーで点検すると共に、使用禁止文字、入力内容について別途慎重に点検する必要がある。

## 2. 添付文書情報XMLファイルの実際

添付文書XMLファイルの具体的な作成方法はS K Wサイトに掲載の「医療用医薬品添付文書XMLコンテンツ利用方針」を参照すること。