

Web 配布文書における XML と HTML の比較

-医療情報を例として-

955150 三井慎也

(指導教員 速水治夫 教授)

1. まえがき

現在，Web 上のほとんどの文章は HTML(Hyper Text Markup Language)で記述され，配信されている。しかし，HTMLは表示するものとして作られたものなのでコンピュータによる意味内容に基づいた処理ができない。Web が大規模になりコンピュータに社会が依存されていく中，限界を感じるさせるものであった。これを打破するために制定されたのが XML(eXtensible Markup Language)である。本研究ではこの二つを比較し，HTMLでは困難な処理を行うことでその差を明確化することを目的とする。

2. XML

XMLは独自の意味内容を持たせるタグ（要素タグ）を作る事ができるマークアップ言語である。一例を以下に示す。

```
<名前>三井慎也</名前>
<年齢>22</年齢>
<アレルギー>ペニシリン</アレルギー>
```

この例のように，HTMLではレイアウト情報しか示さなかったタグに，そのタグに挟まれた文字列についての情報を書くことが出来る。これによりコンピュータにも文字列の意味を理解し，その意味内容に従った処理が可能となる。

3. 要素タグ検出プログラムの作成

本研究では XML の有用性を検証するために特定の要素と内容データとの組み合わせを含む文書を検出するプログラムを作成し，評価実

験を行った。評価用のデータとして医療データを用いた。例を以下に示す。

```
<注射>ペニシリン</注射>
<アレルギー>ペニシリン</アレルギー>
```

上の例のように注射としてのペニシリンとアレルギーとしてのペニシリンが存在するが，このアレルギーとしてのペニシリンだけを取り出し警告を発するという，要素タグとそのタグに挟まれた文字列との組み合わせに基づいた処理を行うことができる。

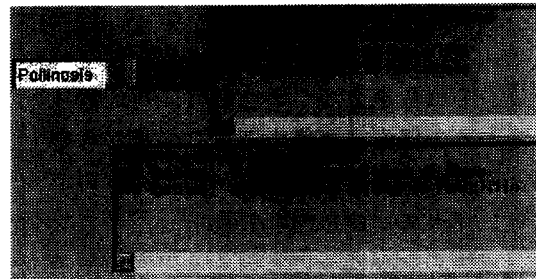


図 1. 実行画面

4. まとめ

対象文章が HTML では困難な処理をするプログラムを作成し，評価実験したことで，XMLの有効性を明確にすることができた。これにより，XMLは Web 上における配布文章をただ単に人間が読むものだけでなく，コンピュータで意味内容に従った処理を可能にする言語であるという事が確認できた。そして，XMLはこれからの Web の利用形態をさらに発展させるものであるということを認識できた。