

# 明確な文章構造の HTML 作成支援

東京工科大学 工学部 情報通信工学科  
坪川研究室 00D066 鈴木 陽平

## 1 背景

インターネットには文字による情報が溢れている。そのほとんどが古い書式の Hyper Text Markup Language (以下 HTML) で書かれている。古い書式の HTML は文章構造と視覚制御情報が混在している。

そのため、文章構造が特定しづらく、要点の抽出、強調する箇所の判別が困難になっている。これは、検索エンジンや音声読み上げ Web ブラウザなどのソフトウェアの開発の障害となっている。

また、視覚制御情報がパソコン向けのものがほとんどであるため、携帯端末などでの閲覧が難しくなっている。

しかし、文書を制作する側も環境が整っていないため、古い書式の HTML を書き続けている。

## 2 目的

HTML から視覚制御情報を切り離し、文書構造を明確にする方法を模索する。また、その方法を実現するため、文章構造と視覚制御情報を別々に編集できるようなソフトウェアの開発を行う。

## 3 文章構造の記述

文章構造の記述には、XHTML Basic1.0 を用いる。XHTML は、XML で定義し直された HTML であり、プログラムで操作しやすいといった XML の特徴をそのまま引き継いでいる。現在パソコン向けに広く使われている HTML4.01 の直接の後継は XHTML1.0 である。XHTML Basic1.0 は、携帯端末などの限られた環境でも閲覧可能のように、XHTML1.0 の機能を一部省略したものである。しかし、ほとんどの文書に対して、機能は十分である。

本研究では HTML の代わりに、XHTML Basic1.0 を用いて、XHTML2.0 で提唱されている文章構造を実現する。現在、2.0 は仕様策定中であり、対応した Web ブラウザもない。XHTML Basic1.0 で 2.0 の記述を再現することにより、2.0 の互換性の低さを解消しつつ、文章構造を明確に記述できる。

## 4 視覚制御情報の記述

視覚制御情報の記述に、CSS を利用する。これにより、HTML と視覚制御情報を分離することが可能である。また、パソコン用、携帯端末用、印刷用など環境ごとに CSS を用意することにより、環境ごとに視覚を制御することが可能である。さらに、複数の HTML が同一の CSS を参照することにより、視覚を統一することが可能である。HTML と CSS の関係を図 1 に示す。

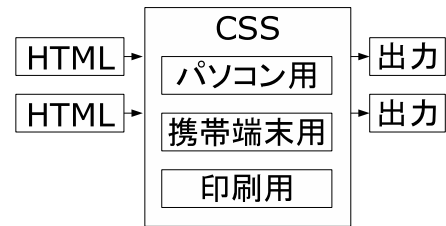


図 1: HTML と CSS の関係

## 5 HTML 作成支援ソフトウェア

開発には C/C++ 言語を用いて、WideStudio、libxml2 で行った。(図 2) HTML を作成する流れを図 3 に示す。作成したソフトウェアのメインウィンドウでは、HTML をツリー表示する。(図 4)

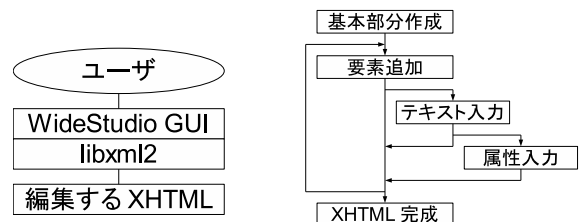


図 2: システム構成図

図 3: HTML 作成の流れ

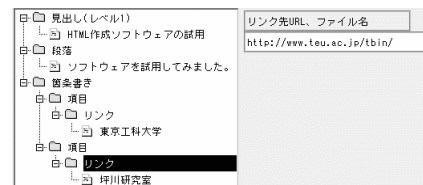


図 4: HTML のツリー表示

## 6 まとめ

HTML から視覚制御情報を切り離すための HTML 作成支援ソフトウェアを開発することができた。しかし、数人に試用してもらったところ、いくつかの問題がわかった。

まず、既存のソフトウェアとの違いが大きいことがあげられる。HTML を作成するコンセプトが違うので、当然のことではあるが、本研究で開発したソフトウェアはやや取っつきにくく、初心者が利用するのは難しい。また、文章構造と視覚制御情報を切り離すことを理解してもらえないこともあった。

開発する予定であった CSS 作成支援ソフトウェアを開発できなかった点も、問題として残った。HTML 作成支援ソフトウェアだけでは、実用的な Web コンテンツの作成は難しい。