

テンプレートを用いたコンテンツ変換

田枝 覚[†] 中川 裕志[‡]

東京大学 情報基盤センター^{†‡}

1 はじめに

携帯端末は、メモリや表示領域の制約から PC 向けの Web ページをそのまま利用することは難しい。従ってコンテンツ変換の需要は非常に大きくなっている。

2 コンテンツ変換システムの提案

2.1 変換における問題点

(1). マークアップ言語の機能の違い

PC 向けの HTML ではテーブルタグがあるが、例えば C-HTML ではそれがない。このとき、単にテーブルタグをはずし、列や行を改行に変換すると、携帯端末では、データ内容が理解し難くなる。

(2). 表示項目を指定すること

携帯端末画面では、PC 向け画面では表示させていたテキストボックスのタイトルを省きたい場合がある。しかしシステムが“タイトル”を判断することは難しい。

(3). 表示順序に関すること

表示領域の制約から、ボタン等を PC 画面とは異なる配置にしたい場合がある。これを自動的に行った結果、意図したものでないとき、それを調整できる方法を用意しておく必要がある。

(4). ページ分割

PC 上 1 ページのものも、データ容量のため、携帯向けでは複数ページになることがある。そのとき、内部リンクのあるものは href の記述を書き換える必要が起こる。

2.2 変換システムの提案

以上のような問題点に対処するために本稿では次のようなシステムを提案する。

システム構成は大きく分けて 1) データ取得部分、2) 携帯端末画面へのデータ再配置部分、3) マークアップ言語変換部分、からなる。1)、2) は後述する“テンプレート”を用いてデータを扱っている。3) はこのシステムに実装されたスクリプト言語で記述する。テンプレート、スク

リプトでの変換法則は、コンテンツと端末種別で変更可能なので、細かなカスタマイズが可能である。

3 変換用テンプレートファイル

システムでは、Web ページから利用したいデータを特定するためのテンプレート、そしてそれらを再配置するためのテンプレートを用意する。それぞれを Template D、Template F と呼ぶことにする。Template D も Template F もシステム認識部分は、HTML のコメントに埋め込まれた形で記述される。コメント形式で記述された様々な指示を、システムの“テンプレート記述言語エンジン”が認識する。

3.1 Template D

Template D は、取得したいデータを特定するためのものである。その目印はコメント形式なので、一般に流通している HTML エディタで画面を見ながら編集でき、画面にも影響しない利点がある。Template D の例は下記ようになる。

Template D 例 (図 1 Template D の①の部分)

```
<td><!-- tuf:data start id="inputKeyword" --><input
type="text" name="keyword" value=""
style="width:180px" size="12"><!-- tuf:data end
id="inputKeyword" -->
</td>
```

□ データ指定

● データ指定 (基本)

これは、繰り返し部分のないデータを指定したいときに用いる書式である。

書式：

```
<!-- tuf:data start id="idname" -->
```

取得したい部分

```
<!-- tuf:data end -->
```

● データ指定 (繰り返しデータの場合)

検索結果画面のように、決まった形式のデータが出力される場合を考える (図 1 Template D の②の部分)。この検索結果ページは、検索結果の構造に変化があるが、外側の“”と“”で囲まれた部分が繰り返されるデータとなっている。このようなデータを取得する場合の指定方法は次のようになる。

```
<!-- tuf:repeat start id="idname" repeat_tag="li/li" -->
```

Transformations of Web contents with templates

[†] Satoru Taeda: Infomation Technology Center, The University of Tokyo

[‡] Hirosh Nakagawa: Infomation Technology Center, The University of Tokyo

サンプルデータ

```
<!-- tuf:repeat end id="idname" -->
```

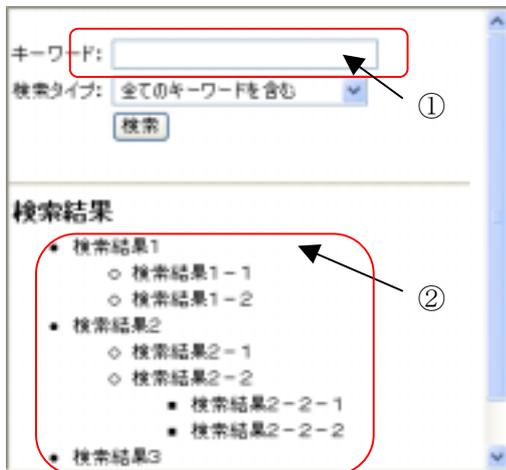


図 1 Template D

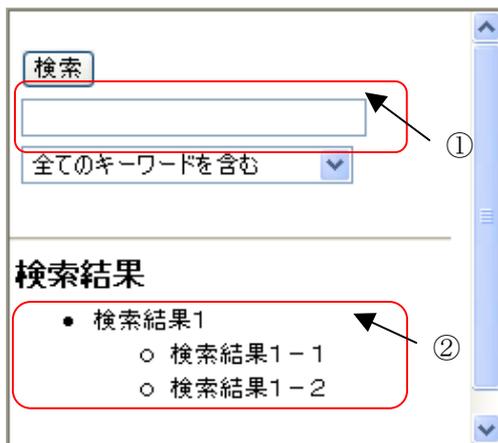


図 2 Template F

□ システムの動作

携帯端末から要求のあったページを取得し、このページを、URL と端末種類から決定した Template D を用い、タグの出現順を比較し目的のデータを取得していく。

3.2 Template F

Template F は出力時のデータの位置関係を指定する役割を持つ。Template F 作成例を図 2 Template F の①の部分に示す。

```
<!-- tuf:data start id="inputKeyword" -->  
<input type="text" name="keyword" value=""  
style="width:180px" size="12">  
<!-- tuf:data end id="inputKeyword" -->
```

画面内の部品は Template D での部品を当てはめ

る形をとる。また上記のコードでは省略しているが、図 2 Template F の②の部分の繰り返しデータもやはり Template D の記述と同じものとなる。

先にも述べたように Template F は表示順序を与える役割を持っている。従って HTML への変換の際は画面表示は Template F とは異なったものになるが、与えられた順序で指定されたデータが表示される。

□ システムの動作

Template F の idname を手がかりにして、対応するデータを当てはめていく。

4 マークアップ言語の変換規則

マークアップ言語の変換は、このシステム独自のスクリプト言語で記述する。変換規則を外部スクリプトファイルにすることで、ユーザ独自の変換規則に変更することが可能になる。システムは先に説明したテンプレートによって加工されたデータを、これらの変換規則に基づきタグ変換し、またファイル分割が必要になる場合ファイル分割を行う。そして、それらに伴うデータの調整を行う。

5 展望

標準的な変換規則を記述したスクリプトファイルはシステム提供時に用意しておいて、適宜そのスクリプトを書き換えるという運用を考えている。今回のシステムでは、同一の Web ページの内部構造がダイナミックに変化する場合、テンプレートが機能しなくなるという弱点があるが、その変化が検索結果ページのようにパターンのあるものならこれまで見てきたように対応している。

6 謝辞

本研究は、東京大学・研究支援推進費、TAO 受託研究「モバイル環境における自然言語処理に関する研究」による援助を受けて行われた。

参考文献

- [1] 中川裕志, “携帯端末向けコンテンツ記述” 言語処理学会第 8 回年次大会ワークショップ「社会情報基盤のための言語・メディア処理」論文集, pp. 33-40, 2002
- [2] 渡部聡彦, 中川裕志 (2002) “多種の表示デバイスへ適応可能なコンテンツ中間表現形式の提案”, 情報処理学会研究会報告, 2002-FI-66-9, 2002