

Table タグによる Web ブラウザでの言語処理結果の汎用表示ツール

中澤 敏明 柴田 知秀 黒橋 禎夫
京都大学大学院情報学研究科

{nakazawa, shibata, kuro}@nlp.kuee.kyoto-u.ac.jp

1 背景

これまでの自然言語処理 (NLP) 研究では、マシンパワーや利用可能なリソースなどの問題で、研究の対象がそれほど大規模ではなく、また処理の複雑さも問題視されるようなことはなかった。しかし近年の NLP 技術の目覚ましい発展や、マシンパワーの向上、さらには Web などに代表されるような大規模リソースへの容易なアクセスが可能となったことにより、研究対象がより複雑化、大規模化している。必然的に、それらを正確に扱う処理 (プログラム) も複雑化、大規模化している。

研究においては、あるシステムで実験を行い、得られた実験結果を議論し、システムへのフィードバックを与えるというプロセスが必須であるが、システムの複雑化により、実験結果を議論する際の手間が無視できないぐらいに増えてきている。つまり、解析ミスなどの言語現象の原因が、処理の複雑化により、複合的なものになる場合や、リソースの大規模化により、そのリソース内の部分を検索するのに時間もかかる場合がある。このような理由から、作業や研究の効率化のために、システムの解析結果を視覚的・直感的にわかりやすい形で表示できるようなツールの必要性を感じ、その構築を試みた。

この際、ツールを深く作りこんでしまい、あるシステムに特化したようなものにしてしまうと、ほかのシステムでの利用が困難となり、その都度ツールを作りなおさねばならず、本末転倒になってしまう。そこでツール自体には必要最低限、かつ十分な機能のみを搭載したシンプルなものにとどめておき、柔軟性かつ頑健性を持たせることを考えた。

さらに、誰もが容易にアクセスでき、多人数での議論においても有用なものにするため、CGI で実現し、Web ブラウザでアクセスすることを考えた。Web ブラウザはどんな種類のマシンにも搭載されており、プ

ラットフォームを気にする必要がないため、十分な汎用性・簡便性を確保できる。また、HTML の機能を用いることにより、視覚的にもわかりやすい表示を実現可能である。

Web 上で視覚的なわかりやすさ、簡便さを得るために、我々は HTML タグである Table(表) を利用することを考えた。これは複数の表を組み合わせるといって単純な表現形式を用いるだけでも、多種多様な情報をわかりやすく表示可能であると考えたためである。この形式を利用すれば、どのテーブルのどのセルに何を表示するのかを指示することによって、様々な種類の解析結果を表示することができる。このとき、表示ツール自体は、指示された内容を表示するだけであることが重要であり、これによって、表示させたい内容によってツールを作り替える必要は一切不要である。表示内容は各システムが、それぞれ知りたい情報を出力すればよいのである。

2 スペック

テーブル表示ツール (以後 TableDisplay と呼ぶ) 自体は各セルごとに指定された内容を持つ、複数のテーブルを表示する機能しか備えていない。各セルに表示する内容やスタイルは、アプリケーション側が指定することになる。

以下に挙げる仕様に従って記述された表示指定ファイルを作成し、TableDisplay に渡すと、指示通りのテーブルが表示される。

2.1 セルの指定

セルは、3つの番号のペアで指定する。1つ目はテーブルの番号、2つ目は行番号 (縦位置)、3つ目は列番号 (横位置) である。命令行であることを示す % の後にこの 3つの数字を置いて、セルの指定をする。例えば、

%% 1 2 3
ここにセルに表示したい内容を書きます。

という指定は、1つ目のテーブルの2行3列のセルに、“ここにセルに表示したい内容を書きます。”と書く。命令行から次の命令行までの内容が、セルの内容となる。ただし、行間にある改行は無視されるため、改行したいときは明示的に
を挿入する必要がある。

また、セル内の文字の位置や、セルの背景色などを指定したい場合は、

%% 1 2 3 valign=top bgcolor=red
この後に
セルに表示したい内容を書きます。

のように、セル指定数字の後にスペース区切りで、プロパティと値を=でつないで指定することができる。これらのプロパティはHTMLに準拠する。上の例では、縦位置を上詰めにし、背景を赤にする。

テーブル番号は自然数だけでなく小数値でも指定可能であり、例えばテーブル番号1.5はテーブル番号1と2の間に表示される。

2.2 テーブルフィーチャ

テーブルにフィーチャを指定することもできる。

%% 2 width=80%

この指定では、2つ目のテーブルの幅が、ブラウザの幅の80%となる。

2.3 ページフィーチャ

ページタイトルなどのページ全体のフィーチャは、%%の直後に指定する(数字をつけない)。

%% title=表示ツール

この指定では、ページのタイトル(<HEAD>内の<TITLE>タグに相当)を“表示ツール”にする。

2.4 リンク

HTMLの特徴的な性質の一つにリンクがある。通常のHTMLでは<A>タグなどを利用して実現するが、TableDisplayも同様の機能を備えている。TableDisplayのリンクには以下の3種類がある。



図 1: 表の作成例

別の表を表示

リンク先も TableDisplay 形式のファイルの場合には、セルのフィーチャに a=file という形式で指定する。

%% 1 2 3 a=file2.txt
file2.txt を開きます

通常の外部リンク

TableDisplay 形式ではなく、一般的な HTML やほかの CGI にリンクしたいときは、extern=URL とする。

%% 1 2 3 extern=http://xxx.xxx.xxx/
外部のファイルにリンクします

追加情報の表示

テーブル内に表示すると複雑になりすぎてしまう情報は、リンクとして別のウインドウで表示させることができる。リンク先に表示させる内容は別ファイルで用意する必要はなく、同じファイル内に埋め込むことができる。

リンク元のセルのフィーチャに label=LABEL1 と、ラベルフィーチャを指定し、以下のように内容を記述する。

%% label=LABEL1
リンク先に表示したい内容

3 作成例

例として図1のような構文解析結果出力を得ることを考える。これには前章のフォーマットにしたがい、以下のように記述すればよい。

またセル 1,4,2 のように、同じセルの内容を複数回記述することも可能であり、様々な処理結果を 1 つにまとめて表示し、詳細なログは LABEL で表示する

なお TableDisplay は京都大学黒橋研究室のサイトからダウンロード可能で、自由に利用することが可能である。¹

- 211 -

