

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 穎夫

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も時間がかかる場合がある。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ふ。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)を利用する考えた。これは複数の表を組み合わせるという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て、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繁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。

```

%% title=KNP 解析結果
%% 1 1 1 colspan=2
# S-ID:1 KNP:2007/02/14
%% 1 2 1 LABEL=1_1b align=right
今日は—— &nbsp;&nbsp;
%% 1 3 1 LABEL=1_2b align=right
天気が—— &nbsp;&nbsp;
%% 1 4 1 LABEL=1_3b align=right
いい
%% LABEL=1_1b
%% LABEL=1_2b
%% LABEL=1_3b
%% 1 3 2 LABEL=1_2d
*
%% 1 4 2 LABEL=1_3d
*
%% 1 4 2
&nbsp;[ガ:天気]&nbsp;
%% 1 4 2
&nbsp;[時間:今日]&nbsp;
%% 1 2 2
&nbsp;DATE:今日&nbsp;
%% LABEL=1_3d
%% LABEL=1_2d

```

1行目でページ全体のタイトルを指定しており、2行目以降で実際のセルの内容を記述している。また LABEL の内容はここでは省略しているが、LABEL が指定されているセルをクリックすることにより、LABEL の内容が別ウインドウで表示される。テーブル 1 の 1,1 のセルには colspan が指定されており、表の 1 行目は 2 つのセルが結合されているのがわかる。

2 行目以降は左側に依存構造木で表示された構文解析結果が、右側には詳細な解析結果が表示されている。ここで注目したいのは、表に表示させたい情報はその表示順に記述する必要はないということである。プログラム処理のログを眺めると、たいていはログの最初に入力に関する部分があり、途中に処理経過があって、最後に出力がある。しかし研究する上で入力と出力を見比べたいことは非常に多くあるが、膨大なログの最初と最後を行ったり来たりするのは面倒である。

しかしこのフォーマットを利用すれば、入力に関する部分を 1 つ目のテーブルに、出力に関する部分を 2 つ目のテーブルに表示し、その他のログは 3 つ目以降に表示するなど、処理結果の表示方法を柔軟にコントロールすることが可能となる。これは単純に表示したい順にテーブル番号を付け変えることで実現される。またテーブル番号は小数値での指定も可能であるため、新しいテーブルを既存のテーブルの間に挿入することも容易である。

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

ことができるなど、解析結果の階層的な表現も可能である。

4 使用例

TableDisplay の実際の使用例を図 2 に示す。ここでは用例ベース機械翻訳システムでの翻訳、および対訳文アライメント結果表示を例に挙げる。用例ベース機械翻訳では入力文と出力文、正解文、スコアなどの情報がまず最初に表示され、次に複数の用例が組み合わされた最終的な翻訳結果が表示されるのが望ましく、図 2 の上の例からこれが実現されていることがわかる。また翻訳に使用された個々の用例は翻訳結果の後ろに一つずつのテーブルで表示されている。

個々の用例には（アライメント結果を表示）というリンクがあり、ここをクリックすると図 2 の下のようなアライメント結果が別のウインドウで表示される。アライメントにおいては、対訳辞書や数字のマッピング、Transliteration など様々な手がかりを用いており、さらに各対訳候補に対して採用、棄却の選択を行なう。アライメントの結果を解析する際に、これらの情報を逐一追っていくのは繁雑になりすぎてわかりにくいし、時間もかかるが、TableDisplay を利用することにより視覚的に分かりやすく表示できていることがわかる。セルの背景の色は、その対応がどの手がかりによって発見されたかを表しており、■がマークされたセルが採用されたフレーズ間の対応となっている。

5 結論

本論文では研究における実験結果の検討の際の補助ツールとして有効である TableDisplay について述べた。TableDisplay は Web ブラウザを用いて視覚的な表示が可能であるが、その機能自体は HTML の<table>タグを利用して複数の表を表示するだけのシンプルな設計となっている。しかし複数の表の組み合わせだけでも様々な情報をわかりやすく、さらには階層的に表示することが可能であり、実験結果の検討効率の向上に大いに役立つことがわかる。

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

¹<http://nlp.kuee.kyoto-u.ac.jp/> にアクセスし、NLP リソース内の解析結果汎用表示ツール TableDisplay からダウンロード可能

機械翻訳システム - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H) リンク » ホーム フォルダ

翻訳結果 (JST_TEST_000350)

入力文:	核蛋白がRNAに結合することによりスプライシング促進機能が得られる事を示唆した。
出力文:	The splicing promotion function was obtained by the combination of the nucleoprotein in RNA.
言語モデル:	The splicing promotion function was obtained by the combination of the nucleoprotein in RNA.
正解 (ref01):	We indicated that the splicing promotion function is obtained by the combination of the nucleoprotein in RNA.
翻訳時間:	5分58秒
翻訳時間:	6分09秒

Cumulative N-gram scoring		Individual N-gram scoring															
1-gram	2-gram	3-gram	4-gram	5-gram	6-gram	7-gram	8-gram	9-gram	1-gram	2-gram	3-gram	4-gram	5-gram	6-gram	7-gram	8-gram	9-gram
NIST: 2.9480	3.2600	3.2600	3.2600	3.2600	3.2600	3.2600	3.2600	3.2600	NIST: 2.9480	0.3120	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BLEU: 0.7495	0.7154	0.6767	0.6316	0.5779	0.5334	0.4919	0.4475	0.3915	BLEU: 0.9286	0.8462	0.7500	0.6364	0.5000	0.4444	0.3750	0.2857	0.1667

翻訳結果

1	└ [3] 核	└ [6] splicing
2	└ [3] 蛋	└ [6] promotion
3	└ [3] 日が	└ [5] the function
4	└ [4] RNAに	└ [1] was obtained
5	└ [2] 結合することにより	└ [2] by the combination
6	└ [6] スプライシング	└ [3] of the nucleoprotein
7	└ [6] 促進	└ [4] in RNA
8	└ [5] 機能が	
9	└ [1] 得られる事を	
10	[1] 示唆した。	

JST TRAIN 112880-1 5.314 (アラインメント結果を表示)

1	└ [0] 効果が	└ * [0] the effect
2	└ [4] 得られる事を	└ [4] was obtained
3	[4] 示唆した。	

JST TRAIN 081848-530 2.214 (アラインメント結果を表示)

1	└ * [6] 糖	└ * [1] it
2	└ * [6] 蛋	* [1] becomes
3	└ @ [6] 日が	└ [10] by the combination
4	└ [10] 結合することにより	└ * [6] of the glycoprotein
5	@ [1] なる。	

add post child: <was obtained> <by the combination>

JST TRAIN 660719-10 7.888 (アラインメント結果を表示)

1	└ [3] 核	└ * [2] reduced
2	└ [3] 蛋	* [2] in the state
3	└ [3] 日を	└ [3] of the nucleoprotein
4	* [2] 還元した場合に	

replace post child: <by the combination> <of the glycoprotein> <of the nucleoprotein>

bond: of the nucleoprotein <- of the glycoprotein, 0, 1

ページが表示されました ローカル インターネット

アラインメント結果 (JST_TRAIN_112880) - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H) リンク » ホーム フォルダ

アラインメント結果 (JST_TRAIN_112880)

Source: b FGF製剤および人工真皮の併用により有効な創傷治癒効果が得られる事を示唆した。
Target: The useful wound healing effect was obtained by bFGF formulation combined with the artificial corium.
アラインメントログを表示
文字アラインメントを表示

Dictionary Num Normalize Transliterare FlexMatch Substr Recover Person Root Match Fill Between

ベース結果を表示

ページが表示されました ローカル インターネット

図 2: 翻訳、アラインメントの結果表示例