

キリルモンゴル語 Web ページの縦書きモンゴル語への自動変換システム*

高紅霞 (学籍番号 200721566)

研究指導教員 : 阪口哲男

副研究指導教員 : 杉本重雄

1. はじめに

縦書きモンゴル語の書き方の特殊性のため電子化が遅れており、Web では縦書きモンゴル語で提供されている情報が少ない。一方、キリルモンゴル語の Web ページは多量に発信されている。そこで、本研究では縦書きモンゴル語圏の人々により多くの Web 情報を利用できるように、キリルモンゴル語 Web ページの縦書きモンゴル語への自動変換システムを開発する。

2. モンゴル語 Web ページとその閲覧

モンゴル語には主に二つの種類がある。一つは縦書きモンゴル語で、もう一つはキリルモンゴル語である。現在、キリルモンゴル語は主にモンゴル国で使用され、縦書きモンゴル語は主に中国の内モンゴル自治区で使われている。

キリルモンゴル語は左から右へ横書きで記述され、行は上から下へ進む。一方、縦書きモンゴル語は上から下への縦書きであって、行は左から右へ進む。従来の縦書きモンゴル語の表示には、ソフトウェア毎に独自に文字コードを定義しているため互換性に乏しい。Windows Vista は Unicode の縦書きモンゴル文字に対応している。ただし、特にレイアウトを指定しなければ文字が横倒しで、横書きになる。最近、Windows Vista が縦書きモンゴル語の表示に使われ始めたがまだ普及していない。一方、キリルモンゴル語は、テキストの電子化が進んでいるので Web ページの多くは標準化された文字コードで提供されている。

縦書きモンゴル語の Web ページの数は非常に少ないがキリルモンゴル語の Web ページは多量に

ある。ただし、内モンゴル自治区では、ほとんどの人々はキリルモンゴル語の読み書きができない。そこで本研究では、モンゴル語の Web ページの閲覧を支援するためにキリルモンゴル語 Web ページの縦書きモンゴル語へ変換システムを開発する。

3. キリルモンゴル語から縦書きモンゴル語への文字変換

本研究では、キリルモンゴル語から縦書きモンゴル語への変換を以下のように行っている。

「伝統的モンゴル語と現代モンゴル語を対象とした双方向的な翻字手法」[1]というキリルモンゴル語から縦書きモンゴル語への変換における先行研究がある。本研究では、[1]で提案したキリルモンゴル語から縦書きモンゴル語への変換手法を利用する。

また、Web からキリルモンゴル語と縦書きモンゴル語の単語の対照辞書[2]を、[1]で解決できなかった不規則な語や固有名詞の問題を解決し、翻字精度を上げるために利用する。

それらに加えて、キリルモンゴル語から縦書きモンゴル語へ変換する際の複数形の表現問題の解決のために表 1 の対応表を定義した。

キリルモンゴル語から縦書きモンゴル語へ変換

表 1 キリルモンゴル語と縦書きモンゴル語の複数形の対応

| キリルモンゴル語の複数形 | 縦書きモンゴル語の複数形 |
|------------------|------------------------------|
| ~ и у л д | ~ᠠᠨᠠᠭ, ~ᠠᠨᠠᠭᠠ |
| ~ и у у л | ~ᠠᠨᠠᠭ |
| ~ {й, н, л, р} д | ~ᠠᠨᠠᠭ, ~ᠠᠨᠠᠭᠠ |
| ~ й с | ~ᠠᠨᠠᠭ, ~ᠠᠨᠠᠭᠠ |
| ~ н у л д | ~ᠠᠨᠠᠭ, ~ᠠᠨᠠᠭᠠᠨᠠᠭ, ~ᠠᠨᠠᠭᠠᠨᠠᠭᠠ |

* “A System for Automatic Conversion of Cyrillic Mongolian Web Pages into Vertical Mongolian Script” by Hongxia Gao

する際には、キリルモンゴル語に対して縦書きモンゴル語の変換候補が複数になる場合がある。その場合、文章の中でどの意味で使われているかが分からないと適切な候補を選べないので、本研究ではユーザが複数の候補から適切な候補を選べるようにした。

4. 縦書きモンゴル語へ変換後の Web ページのレイアウト変換

文字変換後の縦書きモンゴル語 Web ページのレイアウトを変換しないと横倒して読みにくく、ユーザにとって不便である。一方、キリルモンゴル語 Web ページの全体のレイアウトが横書きを主体としているので、縦書きに変換するとレイアウトが崩れて、かえって読みにくくなる恐れがある。

文字変換後の縦書きモンゴル語 Web ページのレイアウトを縦書きにするために CSS の writing-mode を HTML 記述に追加する。本研究ではその追加対象となる HTML 要素と付与方式を定めた。ただし、現時点では Web ブラウザの制約のため、縦書きモンゴル語文章の行の進む方向が逆になり、Web ページの画像が右回りに 90 度回転されることがある。

5. 関連研究

同じ言語における文字変換という点では、「漢字かな自動変換機能等を備えたインターネット学習システムの開発」[3]という研究と Web 上の漢字をひらがな文字に変換して表示できる Web ブラウザ「ひらがな・なびい」[4]がある。しかし、[3]と[4]では漢字をかな文字に置き換えてしまうため、文字数が増えてレイアウトが変わり、見づらくなることがあるが、その解決には至っていない。

6. システムの実現

本システムはプログラミング言語 Ruby を用いて開発した。また、Web アプリケーションフレームワーク Ruby on Rails とデータベース管理システム Mysql を用いた。本システムを利用する際は Windows Vista と Internet Explorer を用いる必要がある。

7. システムの評価

本システムを用いて、24件のキリルモンゴル語 Web ページを縦書きモンゴル語へ変換した。その結果、18 件は正常に変換でき、1 件は縦書きモンゴル語に変換されたが、レイアウトは横倒しになっていた。3 件は変換後の Web ページで文字化けが生じた。2 件は縦書きモンゴル語に変換できなかった。誤った文字コードが指示されている Web ページが存在し、それらは正常に変換できないことがわかった。図 1 は正常に変換できた Web ページの例である。



図 1 正常に変換できた Web ページの例

8. おわりに

本システムを使用することにより、縦書きモンゴル語圏の人々はキリルモンゴル語が分からなくても、キリルモンゴル語 Web ページを閲覧でき、より多くの情報を取得できる。残された課題としては、行の進行方向の対応やより見やすいレイアウトの調整が上げられる。

文献

- [1] 満都拉, 藤井敦, 石川徹也. 伝統的モンゴル語と現代モンゴル語を対象とした双方向的な翻字手法. 情報処理学会論文誌, Vol.47, No.8, pp.2733-2744, 2006.
- [2] 中里致元, モンゴル語 電子化計画.
http://texa.human.is.tohoku.ac.jp/~chigen/md_cnt_j.htm
- [3] 榎本聡, 室田真男, 清水康敬. 漢字かな自動変換機能等を備えたインターネット学習システムの開発. 電子情報通信学会論文誌, Vol. J83-D-I, No.3, pp.384-394, 2000
- [4] ひらがな・なびい.
<http://kids.knowledgewing.com/free35>