Bulletin of Toyohashi Sozo University
2011, No. 15, 195–198

中国笑話集を対象とした文字情報検索システムの機能改善

梅　田　貴　士
山　口　　満
島　田　大　助

中国笑話集研究において，原本および和刻本における文字情報のデータベース化，および検索・比較システムが求められている．本稿では，平成21年度に構築したWebシステム（文字検索，比較，訓読文形式出力，対応画像表示）の改善を行った結果について報告する．

キーワード：中国笑話集，文字データベース，検索・比較システム，主要ブラウザ対応

Ⅰ　はじめに

中国笑話集研究において，中国語で記述された文献（以下原本と記す）と日本で刊行された作品（以下和刻本と記す）における文字情報のデータベース化，および文字検索・比較が可能なシステムが求められている．本稿では，平成21年度に構築した中国笑話集『笑林廣記』文字検索システムを改善した結果について報告する．

Ⅱ　システム概要[1]

1．文字情報データベース

処理の流れを図1に示す．まず，Web上で公開されている原本テキスト[2]をTSV（Tab Separated Values）変換プログラムを用いて1文字ずつ分解・抽出する．次に，原本・和刻本のそれぞれについて，文献[3]を参照しながらページ番号などの情報を追加する．最後に，整理されたTSVデータをDBに登録した．

2．文字検索・比較システム

Apache，MySQL，PHP，JavaScriptを用いて，文献中の文字および振り仮名を対象に，指定文字とDB登録文字との単純マッチングによる検索処理を実装した．この際，原本側のデータおよび和刻本側のデータを同時に表示し，比較できるようにした．また，検索語句の前後文字を表示することで，文脈を把握できるようにした．なお，和刻本の検索結果については訓読文形式で表示させるようにした．この様子を図2に示す．

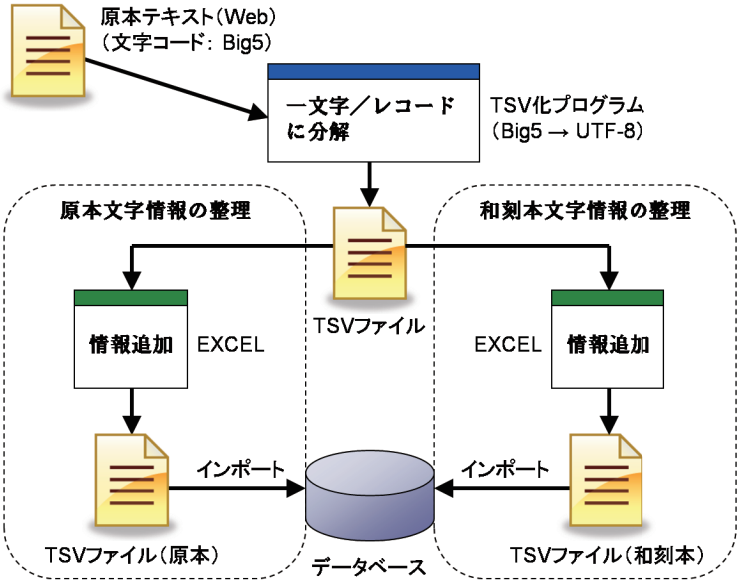

図1　データベース化の流れ

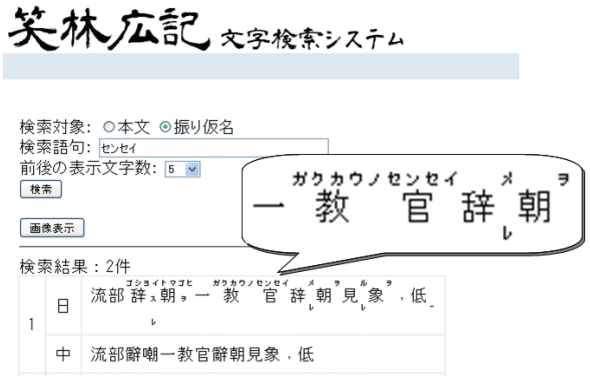

図2　検索結果表示

Ⅲ　前年度からの変更点

前年度に構築したシステムをもとに，機能の改善や追加を行った．

1．主要ブラウザへの対応

豊橋創造大学紀要第14号の報告（以下14号と略す）においては，訓読文の適切な表示を行うためにFirefoxおよびXHTMLルビサポートアドオンが必要であった．今回，ページデザインに用いるCSS（Cascading Style Sheets）の変更と，送り仮名および返り点の記述方

法を変更することでInternet Explorer（IE）やOpera，Google Chromeといった主要ブラウザにも対応した．ただし，IEのバージョン7以前のものについてはやや表示の崩れが発生する．IE（バージョン8）における表示例を図3に示す．また，FirefoxにおいてもXHTMLルビサポートアドオンのない環境に対応した．

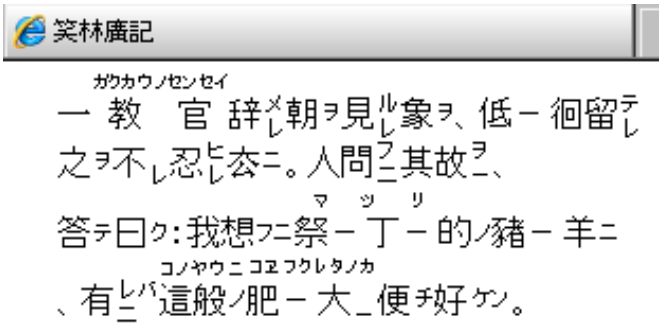

図3　IE（バージョン8）での表示

2．検索方法の追加

14号の方法（図4左）に加え，検索範囲（腐流部や術業部など）を指定しての検索（図4中央），本文全体の表示（図4右）を可能にした．また，検索結果画面を残したまま，異なる検索方法への切り替えを可能にした．

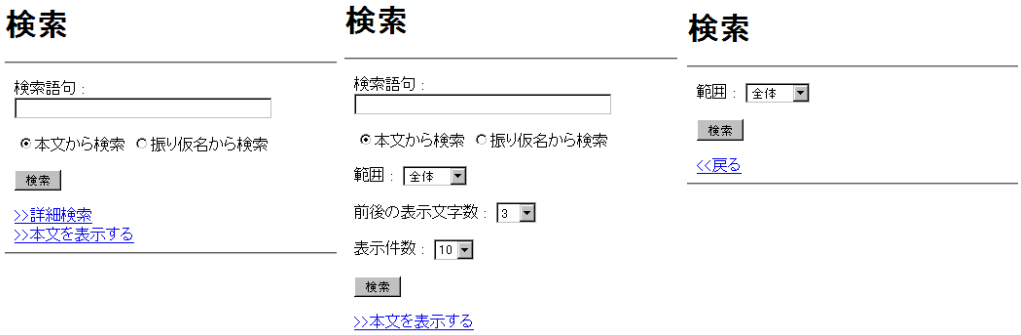

図4　検索フォーム（左：従来，中央：範囲指定検索，右：全文表示）

3．検索結果表示の変更

14号においては，結果画面における表示に，タイトルやサブタイトル，本文の区別なく表示していた．今回，改行と見出しスタイルを適用することで区別できるようにした．さらに検索結果の中で検索語句をわかりやすくするため，検索語句に背景色を付けるようにした．ただし，現状では，検索語句として送り仮名や返り点に含まれる語句（一二点やレ点など）を指定した場合，それらの本文以外の語句についても色がついてしまう問題が残っている．

また，14号では検索の対象を和刻本側のデータのみとしていたが，原本側のデータについても検索対象とし，両者の検索結果を同時に表示するようにした．その様子を図5に示す．

図5　検索結果表示

Ⅳ　まとめと今後の課題

本研究では，中国笑話集研究支援を目的としたシステムの構築および改善を行った．今後は，未整理情報のDB登録，DB登録・修正用インターフェースの実装，異体字など検索語句として文字の入力・指定が困難な場合の検索方法について検討し，よりよいシステムの実現を目指す．

付記

本研究の一部は，平成22年度日本学術振興会科学研究費補助金（基盤研究（C），課題番号21520215）「中国笑話集と日本文学・日本語との関連に関する研究」による支援により行われた．

なお，開発中の文字検索システムは豊橋創造大学内のWebサイトにて公開している（URL：http://document.sozo.ac.jp/cjdb/）．

【参考文献】

［1］　梅田貴士，山口　満，島田大助．「中国笑話集における文字情報のデータベース化」『豊橋創造大学紀要』第14号（2010）pp. 147–150.
［2］『笑林廣記』，http://www.chineselovestory.com/xlgz
［3］　和泉屋金右衛門他板，『訳解笑林廣記』，文政十二年刊（1829），豊橋創造大学附属図書館蔵
豊橋創造大学紀要第14号で『訳解笑林廣記』の刊行年を文政三年（1820）といたしておりましたが，文政十二年（1829）の誤りでした．訂正させて頂きます．