

中国笑話集における文字情報のデータベース化

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、文字検索・比較システムが求められている。本稿では、中国笑話集『笑林広記』について、文字情報のデータベース登録処理、および、同登録情報に基づくWebシステム（文字検索・比較、訓読文形式出力、対応画像表示）の構築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

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）と和刻本（日本で刊行された作品）に対する文字情報のデータベース化、および、文字を検索・比較可能なシステムが求められている。本稿では、中国笑話集『笑林広記』について上記を実現するシステムを検討した結果について報告する。

II 文字情報データベースの作成

原本および和刻本について、本文の文字情報を一文字ずつ抽出し、データベースへ登録する作業を行った。なお、原本テキストとしては、Web上で公開されているものを使用した¹⁾。和刻本としては文献²⁾を使用した。

処理の流れを図1に示す。はじめに、原本テキストを利用してTSV (Tab Separated Values) 化プログラムにより一文字ずつ分解・抽出する。ついで、原本・和刻本のそれぞれについて、文献を参照しながらExcel上で追加情報（文字が出現するページ番号等）を入力する。和刻本については、さらに、訓読のための返り点や送り仮名、ルビ（傍訓）情報等を付加する。図2は、和刻本画像と対応するデータベース登録情報の例である。最後に、整理済みのTSVデータをデータベースにインポートし、文字情報データベースを作成した。

1) 『笑林廣記』, <http://www.chineselovestory.com/xlgz>

2) 『訳解笑林廣記』, 和泉屋金右衛門他板, 文政三年刊 (1820年), 豊橋創造大学附属図書館蔵

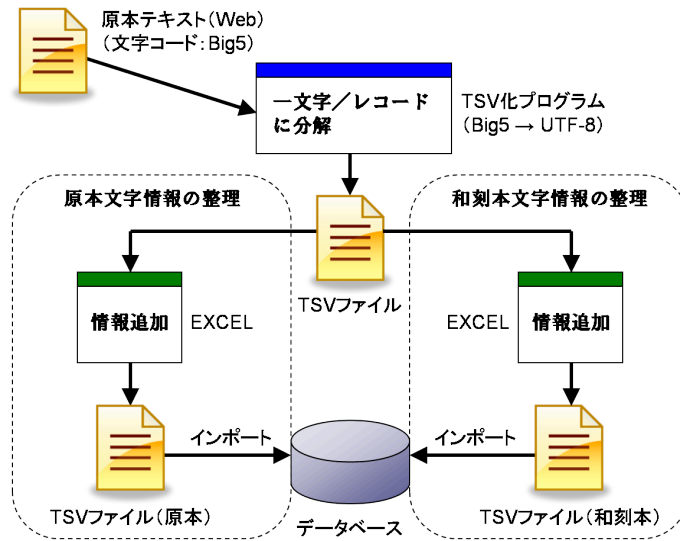


図1 文字情報データベース化の流れ

No	Moji	Kaeri	Okuri	Furi	Furi_cnt
1	腐				0
2	流				0
3	部				0
4	辞	レ	ス	ゴシヨイトマゴヒ	2
5	朝		ヲ		0
6	一				0
7	教			ガクカウノゼンセイ	2
8	官				0
9	辞	レ	メ		0

(a) 和刻本

(b) データベースへの登録情報 (一部)

図2 文字情報のデータベースへの登録

III 文字検索・比較システムの構築

図1の流れで作成したデータベース (DB) を基に、文字検索・比較のためのWebシステムを構築した。なお、構築に際しては、Apache, MySQL, PHPおよびJavaScriptを使用した。

1. 文字検索・比較

文献中の文字およびルビについて、指定文字列と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、原本および和刻本の両者を同時に表示し、比較できるように

した。さらに、検索語句の前後文字を表示し、ヒット位置周辺（文脈）を把握できるようにした。この様子を図3に示す。

2. 訓読文形式表示（整形出力）

和刻本の検索結果については、訓読文形式で表示させる処理を記述した（図3拡大部）。ただし、適切に表示するために、現状ではFirefoxおよびXHTMLルビサポートアドオンを必要とする。

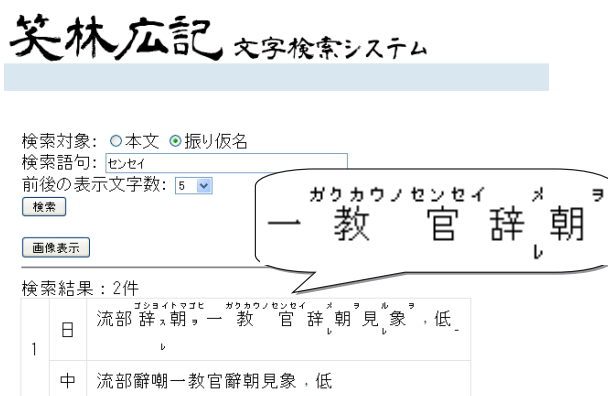


図3 原本・和刻本の検索結果

3. 画像表示

検索結果の文字について、元の文書（画像）における出現位置を表示できるよう処理を記述した。これにより、元の字形を容易に確認することを可能とした。この様子を図4に示す。

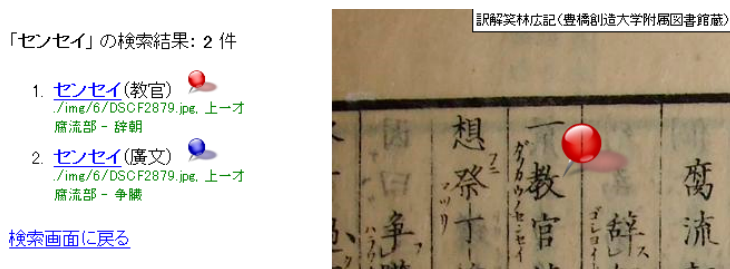


図4 検索結果文字に対応する画像の表示

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を行った。今後は、未整理情報のDB登録作業、ブラウザ非依存ページの記述、および、異体字検索（文字入力による検索が不可能）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21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。

収録データについて

付録として、図1の流れで作成されたExcelデータを取録している。なお、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】

『笑林廣記』，<http://www.chineselovestory.com/xlgz>

『訳解笑林廣記』，和泉屋金右衛門他板，文政三年刊（1820年），豊橋創造大学附属図書館蔵