

19

## 検索語の近接度を限定できる 漢方テキスト複合検索データベースの構築

星野 卓之, 周防 一平, 加畑 聡子, 小田口 浩, 花輪 壽彦

北里大学東洋医学総合研究所

医史学研究や漢方臨床に資する目的で2つの検索語の距離を限定して結果表示ができる漢方テキスト・データベースを開発した。古典研究では、国内文献について写本を含めた異同の多いテキスト間での相互比較が必須であり、引用元として中国書も検討できるとよい。これまで多数のテキストを一括検索するためのデータベースを構築する過程で、インターネットや市販のプログラムなどで流通しているテキストは欠損や文字化けが多いことが判明した。そこで電子テキストを比較検討し改良作成したところ、中国医書約700点、日本医書約200点程度が準備できる見込みとなった。国内文献の写本調査で新たに収集した和書100点あまりを含めると1000点を越えることとなる電子テキスト・データベースを随時更新していくためには、ネットで検索し結果を比較・閲覧できる環境が望ましいと考えられた。

インターネット利用の半数以上がパソコンでなくスマートフォンからといわれる現状に即して、本研究の複合検索比較・閲覧機能はスマートフォンからも参照しやすい環境を優先した。検索窓に入力しやすいよう、専門用語の古字や異体字はスマートフォンでも表示できる文字を優先しつつ共通性が高い文字コードのUnicodeでCJK統合漢字拡張CまたはEまで幅広く表示できるようにした。ダウンロード時に一括して扱えるように、画像などへのリンクを含む書誌情報や目次となる見出し行は、本文の最初にまとめて1つのファイルとなるようにした。

本文表示では見出しとなる行は常時表示させ、そこをクリックすることで続く本文を開閉できるようにした。これにより特殊なプログラムがなくてもブラウザで一階層の疑似アウトライン表示ができ、見出しのみの目次と本文の表示が瞬時に切り替えられるようになった。

検索機能については、多少の異同がある文章を微調整して見つけられるよう、同一行内で2つの検索語間の距離(文字数20字まで)を制限できるようにした。その理由は、既存の検索エンジンにある「あいまい検索」機能は内部処理が不明で、詳細な調査には不相当と考えられたためである。検索結果は、前後の文脈を比較できるよう一行ずつ並べて文字列を色分けし、クリックで該当部の本文を別ウィンドウで表示するようにした。その結果、処方集のように複数語検索に向かない資料についても、2つの構成生薬に絞って処方を探する場合など細かなニーズに応えられるシステムが構築できた。また公開テキストが増えたときのために、ファイル名検索に切り替えられるようにもした。ファイル名はひらがなの書名・医家名と成立年を含み、ひらがなでも簡単に検索できるようにした。2021年2月より北里大学東洋医学総合研究所医史学研究部ホームページで公開し、3月時点では50点程度の資料をアップしているが、今後、3年程度をかけてテキストを増加させていく予定である。データベースの更新情報は北里大学東洋医学総合研究所医史学研究部のFacebookページで発信している。将来的には同時に画像参照も可能なサイトを開発することを検討している。

これまでパソコンのフォルダ内で複合検索できるプログラムは入手可能であったが、検索語間の距離を設定し結果を比較表示できるものは管見に入る限り存在しない。検索語の近接度を限定できる検索システムは、漢方文献に限らず、テキスト・データベースの調査環境を改善するものと考えられる。

(本発表はJSPS科研費JP18K00264の助成を受けたものである)