

# 日本史史料全文テキストデータベースの 構築と漢字入力

永 村 眞

- 
- |                                   |                         |
|-----------------------------------|-------------------------|
| 1. 日本史研究と全文テキストデータベース             | 4. 漢字OCRの導入とシステム開発      |
| 2. 日本史史料全文テキストデータベース構築<br>における問題点 | 5. 全文テキストデータベースの構築とその活用 |
| 3. 問題解決の方向                        | 6. 今後の課題                |
- 

## 論文要旨

日本史研究に全文テキストデータベースが果たす役割はきわめて大きい。その構築にあたっては、漢字字種、漢字字体、不定形の漢字文字列データ、複雑な組版、漢字入力方式、蓄積データの活用方法という側面で、解決すべき様々な問題点が明らかとなっている。そこでその問題点のいくつかについて、その解決を図る可能性を漢字OCRに求めることにした。漢字OCR装置を導入し、日本史史料の自動読取に適合したシステムの開発を試み、具体的なデータベース構築の作業を進めるなかで、全文テキストデータベース構築にこの装置が果たす役割について検討を加えた。その成果として、全文テキストデータベース構築を促進しその活用を図る上での漢字OCRの有効性が確認され、併せて先の問題点の一端に解決の方向が定まるとともに、今後早急に取り組むべき課題の幾つかが明らかになったのである。

なお本稿は、平成2～4年度科学研究費補助金（試験研究B）「日本史史料全文テキスト・データベース構築のための効率的漢字入力システムの開発」（研究代表者：日本女子大学永村眞，研究分担者：タウ技研小篠洋一，東京大学史料編纂所吉田早苗，奈良国立文化財研究所山岸常人）による研究成果の一部である。