

歴史資料を対象とした翻刻支援システムの開発

研究の概要

大阪府河内長野市にある、天野山金剛寺所蔵の漢訳仏典は「金剛寺一切経」と呼ばれ、国内古写経の内容理解において重要なものである。記載内容の分析にあたって、計算機上で利用可能なテキストデータの作成、すなわち「翻刻」が不可欠であるが、経典撮影画像そのものから獲得するのは、人手でも文字認識ソフトウェアを用いるのでも、非常に困難であった。そこで、既存の大正新脩大蔵経テキストデータの中から、ある程度一致するものを基に、撮影画像とテキストとの間で異なる箇所を見つけて編集できるよう、Web アプリケーションを構築した。システムは、必ずしも仏典の専門知識を持たない作業者が使用する「翻刻支援機能」と、その作業結果を専門家が見比べて採否を決める「採用支援機能」で構成される。この分業環境を通じて、各利用者の小さな労力により、質の高い（編集漏れの少ない）経典テキストデータを作成できるようにした。

研究の特徴

各利用者の編集履歴の管理に、主にソフトウェア開発で用いられているバージョン管理ソフトウェアである Subversion を採用した。挿入・削除・置換・交換が考えられるテキストデータの編集において、その情報を効率良く管理するには、従来型のデータベース（DBMS）よりも、テキスト処理や差分といった機能に特化した、バージョン管理ソフトウェアを使用するのが適切であると考えたからである。また編集情報はすべてサーバ側で保持するようにした。これにより、利用者はインターネット接続の PC とブラウザがあれば作業ができ、さらに、複数人の編集作業で考慮すべき衝突の問題も解決した。



図 採用支援機能の画面例

実用化が想定される分野

文書管理, 共同作業

研究者からのメッセージ

データベース (DBMS) だけでなく、バージョン管理, 全文検索, NoSQL などの新技術を積極的に取り入れた情報システムの設計・構築に取り組んでいます。データの保存や活用でお困りでしたら、ご相談ください。

研究分野 : データ工学, デジタルアーカイブ, CSCW

研究者の所属部局・職位・氏名 : 和歌山大学システム工学部 情報学領域・准教授・村川猛彦

本件に関するお問い合わせ : liaison@ml.wakayama-u.ac.jp