

歴史資料を対象とした全文検索システムの開発

研究の概要

寺院に所蔵されている様々な文書（聖教：しょうぎょう）は、作られた当時の人間関係をはじめ、歴史的事象を知る手がかりとなる。そのような研究を支援するため、電子テキストのデータを対象とした全文検索システムを開発した。Webアプリケーションとして実現しており、利用者は、使い慣れたWebブラウザを使用して、GoogleやYahoo!で問い合わせるのと同じように検索語を入力すれば、ただちに該当文書が表示される。検索対象は、文学部の研究者が電子化した約4万件の聖教書誌情報である。



書名	巻数	出典目録	所蔵	書写時期	調べてい	奥書	備考
大日教供養法疏巻下	一帖	築島	石山寺 (校倉聖教一〇函7号)	久寿元年(一一五四)	東大寺点	(奥書) 久寿元年(一一五四)十一月九日於勤修寺書写了/仁平四年(一一五四)六月廿六日於金剛峯寺以他本奉受淨嚴房/阿闍梨御房了 桑門朗龍	月本雅幸氏は淨嚴房は実裡であることを指摘している。

サーバと通信して検索する「クライアント／サーバ型」と、PC（ノートPCやネットブックでもよい）単体で動作し、外部との通信をすることなく検索が行える「スタンドアロン型」の2種類のシステムを構築した。

研究の特徴

聖教書誌情報には文書名などのほか、書籍の奥付に相当する奥書や、研究者によるメモが含まれており、漢文と日本語が混在している。そこで、N-gramに基づくインデックス化が利用可能な全文検索ソフトウェア Hyper Estraier を用いてシステムを構築し、言語に依存せず、瞬時に漏れのない検索ができるようにした。

検索語入力負担を減らすための機能をいくつか実装している。具体的には、「1000」が検索語のときにはこれを西暦年とみなし、「長保二年」という和暦に変換して検索する機能や、新字体で検索しても、旧字体の文書を求めることのできる機能などである。また、検索システムを活用して事前にキーワード（年代、人物名、寺院名、地名）を抽出し、検索結果にリンクを付けており、これにより関連情報の検索がマウス操作のみで行える。

実用化が想定される分野

電子書庫

研究者からのメッセージ

歴史資料に限らず、企業・団体などで作成される情報のデジタルアーカイブ化および全文検索についても取り組んでいる。

研究分野 : 情報検索, デジタルアーカイブ

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