様式５

平成２３年　３月３１日

京都大学人文科学研究所共同研究実績・活動報告書

（　３年計画の　１年度目）

１．研究課題

（和文）情報処理技術は漢字文献からどのような情報を抽出できるか

（英文）What information can be extracted from Kanbun texts with computational methods?

２．研究代表者

（氏名）山崎直樹

３．研究期間

平成　２２年　　７月　から　平成　２５年　　３月　まで

４．研究目的（400字程度）

本課題は、現代の情報処理技術が東アジア古代社会の遺産として残された漢字文献からどのような情報を抽出できるかという問題に対し、複数の角度からその可能性を探り、人文情報学の基礎を築くことを目的とする。ここでいう「人文情報学」とは、文献を機械処理することにより、i) 人手では不可能な大量のデータを扱い、 ii) 人手による処理では帰納できない類の情報を抽出し、 iii) 得られた情報を機械可読なかつ再加工可能な形式で蓄積する学問分野を指す。

本課題では、以下に例として挙げる処理が可能であるかを試みる予定である。a) 白文の古典中国語のテクストを解析して句読点を施す、b) 墓誌銘や地方誌などの定型的なテクストから研究者が望む情報を抽出する、c) 返り点の施された漢文は構文情報をアノテーションしたコーパスであると考え、構文情報ごとコーパス化をする、d) 経典等に付された注釈は元のテクストに付されたメタデータであると考え、それを機械可読な形式で関連づける、e) 定型的な韻文などで常に関連づけられる語句は、ある種のネットワークを成していると考え、それを体系化する、などである。

５．本年度の研究実施状況（400字程度）

2010年7月23日に最初の会合を開催したのち、各情報処理技術に関して、電子メールで意見交換をおこなった。さらに12月3日に二度目の会合を持ち、それらの技術を、今後の研究にどう活かしていくか議論をおこなった。その後、2011年1月21日と2月4日に会合を持ち、モデルによる文献研究に関して議論をおこない、2月18日に公開シンポジウム『文字と非文字のアーカイブズ／モデルを使った文献研究』を開催した。

６．研究成果の概要（400字程度）

今年度はキックオフの年でもあり、とりあえずは各情報処理技術、特に文字と非文字のアーカイブズに関する処理技術のサーベイをおこない、来年度に向けての足がかりとした。文字資料に関しては、これまでのデータベース研究による蓄積がかなり膨大で、検索技術もすこぶる発展しているのに対し、写真や動画についての検索技術は未だほとんど手つかずの状態にあり、これらの技術をどのように構築していくのかがメインの議論となった。なお、公開シンポジウム『文字と非文字のアーカイブズ／モデルを使った文献研究』の予稿を含め、本共同研究の成果情報は、逐次 http://kanji.zinbun.kyoto-u.ac.jp/~ymzknk/kanzi/ で全世界に公開しており、そちらも参照されたい。

７．共同研究会に関連した公表実績（出版、公開シンポジウム、学会分科会、電子媒体など）

2011年2月18日に公開シンポジウム『文字と非文字のアーカイブズ／モデルを使った文献研究』を開催した。さらに、公開シンポジウムの予稿集を、「文字と非文字のアーカイブズ／モデルを使った文献研究」(全国共同利用・共同研究拠点「人文学諸領域の複合的共同研究国際拠点」、京都、2011年2月)として発行した。また、本共同研究の成果情報をWWWサイト http://kanji.zinbun.kyoto-u.ac.jp/~ymzknk/kanzi/ において公表した。