

京都大学人文科学研究所共同研究最終報告書（終了）

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ー 人文情報学の基礎を築く

(英文) What information can be extracted from Kanbun texts with computational methods? -- A contribution to fundamental research in Digital Humanities --

2. 研究代表者

(氏名) 山崎直樹 (関西大学)

3. 研究期間

平成22年7月 から 平成25年3月 まで (2年9ヶ月間)

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最終目的とした。そして、そのために、現状に対する認識を踏まえ、下記の課題(甲)-(丙)を検討をすることを作業目的とした。課題(甲): 非文字データのように、これまでメタデータを付与することが困難などと考えられてきたデータへのメタデータ付与にはどのような問題点があり、どのような可能性があるか。課題(乙): 文字データにせよ、非文字データにせよ、メタデータを付与することにより、単なる機械可読ではなく、機械が意味を読み取ることが可能となったデータは、そのデータ相互をどのように関連づけたらよいのか。「セマンティックウェブ」という枠組みで示された方向でよいのか。課題(丙): 我々が分析の対象とする文字データの集まり=テキストは、どのような構造をもっているものとしてモデル化できるか、また、非文字データの集積体の基本的なデータフォーマットは、画像や動画であると考えられるが、それらの集積体 (例: 記録映画など) はどのような構造をもっているものとしてモデル化できるか。

5. 研究成果の概要 (400字程度)

「研究目的」で述べた課題(甲)(丙)については、静止画像資料や動画資料へのメタデータ付与、文字と画像が混在し線的な展開をしない資料へのメタデータ付与の可能性を探った (詳細については、公開シンポジウムの資料を参照)。また、課題(乙)については、大規模なLinked Open Dataの可能性、国立情報学研究所のCiNiiの可能性、Dublin Coreを拡張した国立国会図書館のDC-NDLなどの可能性について、とくに外部から公開シンポジウムに報告者を招き、その現状と問題点、将来への展望について、理解を共有した。課題(丙)については、「大規模ネットワークの物理的構造」への理解が不可欠であると考え、複数の公開セミナー・公開シンポジウム

を通じて、外部から招いた専門家よりネットワーク科学の基礎を学び、その文献研究への応用可能性を検討した（詳細については、公開セミナー・公開シンポジウムの資料を参照）。

6. 本研究課題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）
 1. 【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（2011年2月18日、於京大人文研本館）
 2. 【公開セミナー】ネットワーク科学は道具箱（2011年11月19日、於京大人文研本館）
 3. 【公開シンポジウム】情報の構造とメタデータ（2012年2月24日、於京大人文研本館）
 4. 【公開シンポジウム】すべてをコンピュータの中に（繋がってしまったデータとその未来）（2013年2月16日、於京大人文研本館）

7. 研究成果公表計画および今後の展開等
 - 上述のシンポジウム・セミナーに関する文字資料は、専用ウェブサイトで閲覧できる（<http://kanji.zinbun.kyoto-u.ac.jp/~ymzknk/kanzi/>）。
 - 上述のシンポジウム・セミナーの録画は、Ustreamで視聴できる(人文:USTREAM <http://ustream.tv/channel/zinbun/>)。

（この報告書は本研究所HPなどで公表されます。）