

立命館大学アート・リサーチセンター
 文部科学省 国際共同利用・共同研究拠点
 「日本文化資源デジタル・アーカイブ国際研究拠点」
 2020年度 国際共同研究成果報告書〔研究費配分型〕

2021年 5月 7日 提出

1. 研究課題名	
芳賀矢一『日本人名辞典』からの歴史人物情報の自動抽出および情報構造化 (英文課題名) Automatic Extracting Personal Information of Historical Persons from "Nihon Jinmei Jiten" and Creating Structured Data	
2. 研究代表者	
氏名 (ふりがな) きむら ふみのり	所属機関・職名
木村 文則	尾道市立大学
3. 研究分担者 (合計: 2名)	
氏名 (ふりがな)	所属機関・職名
前田 亮(まえだ あきら)	立命館大学 情報理工学部・教授
後藤 真(ごとう まこと)	国立歴史民俗博物館・准教授
4. 研究課題の概要 (300字程度) (申請書から変更がある場合は、変更点分かるように明記してください)	
<p>日本の歴史研究を行うにあたり、最も基盤となる情報の一つとして人名があげられる。現在国立国会図書館が人名典拠情報を提供しているが、書籍など限られた人名になっており、歴史上の人物を広く網羅できていない。そこで、約50000人の人名を収録し、かつ著作権が切れている芳賀矢一編『日本人名辞典』(以下、『芳賀人名辞典』)を活用した人名データを作成することで、広く歴史的人名の基礎状況を提供することができる。現状はこの作業は専門家が人手により行っているが、本研究において抽出の自動化を実現することにより、その作業の支援を行うことを目指す。</p> <p>また、このデータは、機械可読であることが情報基盤の最大の条件であるので、Linked Data として情報を記録し、データベースへと格納する。これにより、格納された情報の意味を表現できるとともに、他のデータベースの情報とのリンクが容易となり、日本史をはじめとする諸研究の促進が期待される。</p>	
5. 研究成果の概要 (この項は、本センターのホームページ・紀要等で公開することがあります)	
<p>芳賀矢一編『日本人名辞典』に記述されている人物の解説文を解析することにより、その人物の人物情報を取得した。本研究においては、ニューラルネットワークの一種である word2vec を用いて、抽出すべき属性に該当する単語を検出し、その単語を人物情報として取得した。複数の属性(「役職」「年号」「地理」「別名」)においては、高い抽出精度を達成した(適合率 0.76~1.00)。</p> <p>本手法は改善の余地が残ってはいるが、専門家が人手により人物情報の抽出作業を行う際に、提案手法により支援を行うことができるだけの成果が得られた。</p>	

6. 研究業績 (日本語以外に英語名称もあるものは、できるだけ日英両言語でご記入ください)

- (1) 著書
- (2) 論文
- (3) 研究発表等
- (4) 主催したシンポジウム・研究会等
- (5) その他研究活動 (報道発表や講演会等)
- (6) 受賞学術賞
- (7) 科学研究費助成事業
- (8) 競争的資金等 (科研費を除く)
- (9) その他