書 式 7

助成番号	１６－565

松下幸之助記念財団 研究助成

研究報告

（MS Word データ送信）

【氏名】木村麻衣子

【所属】（助成決定時）慶應義塾大学大学院

【研究題目】漢籍デジタルアーカイブの国際的連携に向けた基礎的研究

【研究の目的】（４００字程度）

わが国には古典籍デジタルアーカイブが分散して存在しており,その規模も様々である。小規模のデジタルアーカイブは,他のデジタルアーカイブと連携しておらず,統合的に検索できない。和古書のデジタルアーカイブについては,国文学研究資料館が「日本語の歴史的典籍の国際共同研究ネットワーク」を構築中であるが,比較的大規模で,画像データおよびメタデータを提供可能な体制を整えた機関のみが連携対象となっている。大規模な連携の枠組みから漏れてしまう小規模機関の古典籍デジタルアーカイブにも,連携の仕組みを用意する必要がある。とりわけ漢籍については,日本国内の漢籍データを連携する枠組みすらない。本研究では,将来的に国内および海外の漢籍デジタルアーカイブを統合的に検索できるようにするための基礎的研究として,漢籍の統一タイトルデータベースの試験的な構築と,連携の実証実験を通じて,統一タイトルを介したデジタルアーカイブ間連携の効果を明らかにすることを目的とする。

【研究の内容・方法】（８００字程度）

① 統一タイトルデータを記録するためのデータベースを構築する。

漢籍著作名典拠データベース KWMA-san を構築し，公開した。(https://zoshoin-db-zosan.herokuapp.com/works)。著作名のタイトルデータは，杜信孚, 王剑编著. 同书异名汇录. 江苏古籍出版社，2000. を元にした。『同书异名汇录』は簡体字表記であるが，漢籍を取り扱う際の基本文字種は繁体字であるので，繁体字での入力を基本としたほか，タイトルについてのみ簡体字，ピンインによる入力を必須とした。そのほかに，日本漢字やハングル，韓国漢字の入力フィールドも設けたが入力は限定的である。異体字検索機能を実装し，基本的には異体字で検索してもヒットするようにした。『同书异名汇录』には，同一著作のさまざまなタイトルの中でどれが最も優先されるべきタイトル（統一形，または典拠形タイトルと呼ばれる）かの明示はない。そこで，全国漢籍データベースの検索結果などから，典拠形タイトルの選定を木村が行った。

② CiNii Books の API を使用して，①に個別資料名を追加する。

著作タイトルは，それぞれの個別の資料の巻頭題とは異なる可能性もあるため，それぞれの著作に属する個別資料を漏れなく検索できるようにするために, 著作データに, それに属する個別資料名を付加する。CiNii Books の API を通じて CiNii Books の書誌データに書かれている書名，別書名を個別資料名として①の著作データにリンクした。

③ 統一タイトルデータを Linked Open Data(LOD)として公開する。

KWMA-san で公開中の 200 件分の著作データおよびこれに紐づく個別資料データについて，LOD で公開した。公開サイト URL は https://harumukanon.github.io/kwma-san/である。

④ 既存の複数の漢籍デジタルアーカイブを用いて，漢籍の書名のみで検索した場合と，著作名典拠データを介して検索した場合の精度や再現率の差を比較する。

CiNii Books, NDL サーチ，および全国漢籍データベースのそれぞれを，a)著作名の典拠形のみで検索した場合，b)著作名の典拠形と異形を OR 検索した場合，c)個別資料名とその別名も②とあわせて OR 検索した場合で精度と再現率を比較した。

【結論・考察】（４００字程度）

3 つのデータベースを，a)b)c)それぞれの検索式検索し，それぞれの精度，再現率，F 値を単純平均した。いずれのデータベースでも a)に比べ b)は精度が低くなるものの，再現率が 90%台まで上昇する。したがって，典拠形のみでなく，典拠形に異形を加えた検索を行うことで，典拠形だけでは検索できなかったレコードにたどり着ける可能性を提供すると言える。

いずれのデータベースでも a)と c)および b)と c)の再現率には差があることから，個別資料名を加えて検索することで適合レコードが増えていると言えるが，同時に精度が大きく下がっている。再現率と精度の調和平均を示す F 値も，CiNii Books を除き a)b)間で低下しているが，b)c)間ではさらに大きく低下した。特に，全国漢籍データベースでの精度の低下が著しい。以上より,検索式 c)は再現率を高めることには一定の効果があると言えるが，精度の低下への対応が必要である。