Advances in Social Sciences 社会科学前沿, 2021, 10(6), 1550-1553
Published Online June 2021 in Hans. http://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2021.106214

Hans汉斯

关于数字化检索与文献学研究的几点思考

杨　超

湖南师范大学历史文化学院，湖南　长沙
Email: yangc837@163.com

收稿日期：2021年5月21日；录用日期：2021年6月16日；发布日期：2021年6月23日

摘　要

数字化时代对文献学研究提出了更高要求，穷尽史料成为可能，然而智能检索尚不能代替阅读分析，文献学的基本方法没有随之发生根本性变化。史料的丰富，要求研究者必须对其背后的典籍、人物乃至文化、社会有深入了解，惟此方能得出可靠的结论。因之，文献学研究的实际难度较前数字化时代增大。

关键词

数字化，检索，文献学，古籍

Some Thoughts on Digital Search and Philology Research

Chao Yang

College of History and Culture, Hunan Normal University, Changsha Hunan
Email: yangc837@163.com

Received: May 21st, 2021; accepted: Jun. 16th, 2021; published: Jun. 23rd, 2021

Abstract

In the digital age, there is a higher demand for philological studies. It is possible to find all the historical materials. However, intelligent retrieval cannot replace the reading and analysis, the basic method of philology did not follow the fundamental changes. The richness of historical data requires the researcher to have a deep understanding of the relevant ancient books, characters and even culture and society, and only then can we draw a reliable conclusion. Thus, the practical difficulty of philological research is more than that of non-digital age.

文章引用: 杨超. 关于数字化检索与文献学研究的几点思考[J]. 社会科学前沿, 2021, 10(6): 1550-1553.
DOI: 10.12677/ass.2021.106214

Keywords

Digitizing, Search, Philology, Ancient Books

1. 引言

当前，古籍数字化事业发展的如火如荼，为古代学术研究提供了极大便利，也在某种意义上，改变着学术研究的理路。然而，从目前数字古籍和古籍数据库的实际应用来看，古籍数字化事业还处于初级阶段，缺乏整体协调，检索分析亦未臻智能化，可供发掘的空间还很大。不过，就目下而言，已有大量史料借助于此进入学者视野，推动了研究的进展。我们仅就目前古籍数据库尚不完备的情况下有关数据检索与文献学研究的几个问题，谈一些浅显的感想，以期大雅方家续发宏论。

2. 数据检索与史源考察

对史料进行史源学的分析考证是历史研究的基础。大数据时代的古籍数字化将会为史源学研究带来极大便利，即使今天，这种便利也已经显现。数字古籍数量的不断增长，尤其各种古籍数据库便利的检索，为史源学研究提供了丰富的材料。过去需要翻检很多书籍才能发现的问题，今天借助检索可能几分钟就发现了，节省了大量手工翻阅功夫的同时，为深入细致的考证提供了更多时间。

史源学的研究不单单是罗列史料，更重要的是追本溯源，深入探究史料本身的复杂构成，这就不仅仅是检索能解决的问题了，还要具备一些文献学的基本知识。因此，古籍数字化对史源学研究的推动是潜在的，如果分析不到位，对于研究并没有太大价值。在史源探讨方面有一个误区，就是以最早出现的记载作为考察对象的直接史料来源，这其实是有问题的。举一个例子。

《资治通鉴》卷一七：“太皇窦太后好黄、老言，不悦儒术。赵绾请毋奏事东宫。窦太后大怒曰：‘此欲复为新垣平邪！’阴求得赵绾、王臧奸利事，以让上。上因废明堂事，诸所兴为皆废。下绾、臧吏，皆自杀。丞相婴、太尉蚡免，申公亦以疾免归。”[1]

经过检索，我们发现此处记载可能的史料来源主要有以下几则：

《史记·孝武本纪》：“会窦太后治黄老言，不好儒术，使人微得赵绾等奸利事，召案绾、臧，绾、臧自杀，诸所兴为者皆废。”[2] 452

《史记·魏其武安侯列传》：“太后好黄老之言，而魏其、武安、赵绾、王臧等务隆推儒术，贬道家言，是以窦太后滋不说魏其等。及建元二年，御史大夫赵绾请无奏事东宫。窦太后大怒，乃罢逐赵绾、王臧等，而免丞相、太尉，以柏至侯许昌为丞相，武强侯庄青翟为御史大夫。”[2] 2843

《史记·儒林列传》：“太皇窦太后好老子言，不说儒术，得赵绾、王臧之过以让上，上因废明堂事，尽下赵绾、王臧吏，后皆自杀。申公亦疾免以归，数年卒。”[2] 3122

《汉书·窦田灌韩传》：“太后好黄老言，而婴、蚡、赵绾等务隆推儒术，贬道家言，是以窦太后滋不说。二年，御史大夫赵绾请毋奏事东宫。窦太后大怒，曰：‘此欲复为新垣平邪！’乃罢逐赵绾、王臧，而免丞相婴、太尉蚡，以柏至侯许昌为丞相，武强侯庄青翟为御史大夫。”[3] 2379

《汉书·儒林传》：“太皇窦太后喜《老子》言，不说儒术，得绾、臧之过，以让上曰：‘此欲复为新垣平也！’上因废明堂事，下绾、臧吏，皆自杀。申公亦病免归，数年卒。”[3] 3608

通过阅读，我们可以发现，《资治通鉴》的这则记载在对《史记》、《汉书》相关材料进行仔细比对、分析的基础上，还原了汉武帝迫于窦太后压力暂时黜抑儒术的来龙去脉，对《史记》、《汉书》含糊不清的地方做了恰当的梳理。它的史料来源不是单一的。其中“赵绾请毋奏事东宫。窦太后大怒曰：‘此欲复为新垣平邪！’”很明显采自于《汉书·窦田灌韩传》，虽然《史记·魏其武安侯列传》也有类似的话。而“下绾、臧吏，皆自杀”，亦是采自《汉书·儒林列传》，虽然《史记·儒林列传》也有“尽下赵绾、王臧吏，后皆自杀”的话。因此，最早的记载未必就是考察对象真正的直接史料来源。面对构成复杂的史料，检索只是节约了我们的时间，并不能代替分析而一劳永逸。

3. 数据检索与文献甄别

对于文献考证而言，首要的是尽可能穷尽史料，大数据时代为实现这一目标提供了条件，当然就整体而言，今天还做不到，不过部分已经接近了，如先秦两汉的传世史料。而魏晋到宋元还在一步步充实，明清任重道远。倘若所有史料均已齐备，如何查考就成为第二个问题。传统的办法是一部一部的阅读，今天越来越多是靠电子检索。电子检索是否能完全代替传统的阅读，答案无疑是否定的。因为不是所有内容都能检索出来。以最简单的人物为例。古人称谓有名、字、号、官名、排行等多种形式，为免遗漏，这些称谓都要进行检索。即便如此，仍有未尽之处。如某公、某郎、某大人这种泛称，倘若检索，信息是海量的，显然不合适。这就要求我们对这种称呼可能出现的著作要熟悉，熟悉来源于传统的阅读积累。目前，由哈佛大学、台湾“中研院”史语所、北京大学正在开发建设的中国历代人物传记资料库(CBDB)有望一键解决古人名、字、号、排行等称谓的检索。但泛称问题仍不能有效解决[4]。即使有一天泛称被收入数据库，也可通过关联一键检索出来，除非人工智能技术取得重大突破，毋庸置疑的是，大量的传统阅读仍是后台编程工作的坚实基础。

传统阅读加上电子检索或可以减少查考的遗漏，接下来就是第三个问题——具体分析了。面对大量的文献，如何分析，考验的实际还是基本功。举一个例子。在电子检索的过程中，我们经常会遇到许多内容大致相同而不断重复出现的文献，在一条一条阅读完后往往会发现有许多文献对我们的考证没有用，因为对文献考证而言，一般应以最早出现的记载为依据，故而后代尤其是较晚时代重复出现的记载就没有太多价值了。如对唐代文献而言，清代的记载大多转抄于前代，故而价值不大。为了节省时间，就需要我们对不同时代的典籍类别及特点有一定了解，这样才可以略过一些文献而只阅读有用的部分，甚至在检索时就忽略掉一些书籍。由此看来，传统的目录学修养在大数据时代更显重要。

4. 数据检索与古籍辑佚

古籍数字化似乎对文献学研究范围内的版本校勘和辑佚冲击最大。版本可以自动比对，确实提高了工作效率和准确性，值得推广，但是版本系统的分析，尤其文字是非论定，还需要人去处理，版本刊刻流传的考察就更不必说了。

辑佚虽然借助数据库检索可以扩大范围，提高效率，但检索时也要具备传统辑佚学的相关知识，而对佚文的分析目前更是电脑代替不了的。就检索而言，至少要注意以下问题。古人引书常引大意，这样很多佚文可能检索不出来；古人引书常不注明出处，对于查找文献来源增加了难度；古人引书常用简称或代称，无形之中增加了检索量。明了这些，才能更好地进行检索。但显而易见的是，单凭电子检索不足以将古书佚文搜罗完备，还必须要有针对性的阅读。而就检索结果进行分析时，就更得依靠传统方法了。仅就形式方面而言，由于古人引书时正文、注文、引文常常混淆，引书时常常署名不清，古书经常

有同书异名和同名异书等现象的存在，就需要运用考证的方法对佚文的归属进行判断，以免误辑。在此基础之上，方可进行内容的考察。查考内容首先要明确古书的体例，其次再进行文字的校勘。但无论怎样，文献学的基本方法还没有因为古籍数字化而发生根本改变。

目前，大大小小的古籍数据库已经开发了不少，但大多存在重复收录、校勘不精、检索不便、缺失尚多等问题，从一定意义上说，还不能满足文献学研究的需要[5]。与我们理想的智能分析系统，还有相当距离。然而，和前数字化时代相比，今天的我们，不仅大大摆脱了奔波各地搜讨文献的辛劳，而且还看到了很多前人无法阅读到的文献，这已经大大便利了相关研究。若能有效利用现有的古籍数字化成果，理论上讲，亦能做出超越前人的成绩。数字化时代对文献学研究提出了更高的要求，穷尽史料成为可能，然而正如上文所例举的三个方面，检索尚不能代替阅读，传统文献学的修养在此时更显重要。有学者指出，大数据时代史学研究碎片化现象严重，对文献学而言，似乎更明显，究其原因，在于我们对文献背后的典籍、人物乃至文化、社会均一知半解，不能形成整体观照，以至于只见树木不见森林，在这种情况下做出的结论是大可商榷的。数字化为我们提供了越来越丰富的文献，这些文献不仅需要智能检索，更需要静下心来仔细阅读，惟此方能得出可靠的结论。

基金项目

本文为湖南省教育厅科学研究项目“五代十国教育史研究”(编号：18A015)阶段性成果。

参考文献

[1] 司马光. 资治通鉴[M]. 北京: 中华书局, 1956: 557-558.

[2] 司马迁. 史记[M]. 北京: 中华书局, 1959.

[3] 班固. 汉书[M]. 北京: 中华书局, 1962.

[4] 潘俊. 面向数字人文的人物分布式语义表示研究——基于 CBDB 数据库和古籍文献[J]. 图书馆杂志, 2020, 39(8): 94-102.

[5] 付艳. 基于内容的古籍检索技术研究[C]//第二届中国古籍数字化国际学术研讨会论文集. 北京: 五洲传播出版社, 2011: 124-130.