

# 数字人文知识生产转型过程中的 困境与突围

陈 静\*

**摘要** 针对一系列关于数字人文的“计算”及“工具”性的争论，本文指出数字人文是数字知识生产转型中的必然路径，但需警惕算法依赖和工具化约，需要突破在乐观和悲观意义上同时有效的“技术黑箱”。如将“计算”问题视为单纯的算法、工具和方法，是无法解决数字人文的“计算”困境的，因此，数字人文需要从知识谱系上进行拓展，尤其需要重建包括媒介研究在内的其他思想谱系的关系，将数字人文放置到“数字文化”的大语境中，突破数字人文的计算困境。

**关键词** 数字人文 知识生产 计算 算法 媒介研究

**Title** *The Crisis and Solution of Digital Humanities in the Transformation of Knowledge Production*

**Abstract** Targeting to the debates on “computation” and “tool” of digital humanities, this essay points out that digital humanities is the unavoidable approach to the transformation of digital knowledge production but needs to be cautious with the dependence on the algorithm and the reduction the digital humanities to tool and break the black box of technology affected in the sense of optimism and pessimism. Regarding the computation as the simple algorithm, tool and method, it is impossible to resolve the crisis of computa-

---

\* 陈静，文学博士，南京大学艺术学院副教授，主要从事媒介文化与数字人文研究。

基金项目：国家社会科学基金重大项目“西方美学经典及其在中国传播接受的比较文献学研究”（17ZDA021）。

tion of digital humanities. Hence, it would be necessary to expend the genealogy of knowledge of digital humanities and reconnect the relationship between digital humanities and other knowledge category, such as media studies. Only resetting the digital humanities in the broader context of digital culture, the crisis of computation would be able to be resolved.

**Key words** Digital Humanities Knowledge production Algorithm Computation Media studies

数字人文一直具有争议性。最近一次颇有影响力的论争是由布伦南 (Timothy Brennan) 在《高等教育纪事报》(The Chronicle Of Higher Education) 网站上发表的文章引起的。他旗帜鲜明地批评数字人文研究中对算法的依赖,使数字人文学者在面对文本时只看到了通过算法所呈现出的文本特点,比如词频,却无法触及文本中有价值的内容部分。他还针对内容的量化分析、文学批评中的“远读”策略和“文学模式识别”等指出,由于数字人文学者们只是看到了表层的数字和数据,不能像使用大脑那样使用计算机深入思考,所以计算机只能回答其所能回答的问题,而无法开展真正深入的批判性反思;同时还因为数字人文学者将“更多信息混淆为更多知识”,所以使数字人文无法在其所施用的方法之外进行反思,获得认识论或者方法论上的价值。因此,数字人文不过是占有少量资源或者长期处于学界边缘的年轻学者的手段,借此在已经划定格局的学术场内争取更多文化资本,获取地位提升。布伦南指出,“与其说数字人文是一场革命,不如说数字人文为了反对主流形式,从而强行将人文从其存在原因中剥离出来的那个楔子”<sup>①</sup>。可见,在布伦南这里,数字人文是全面破产的状态,非但没有从方法上与传统人文研究一样有效(数字人文研究的效果取决于算法的有效性),更因为计算机无法像人脑一样有智慧,无法对自身进行认识论上的反思,也就无法生产知识。

这种强调数字人文政治意义高于其研究价值的观点是过去十多年不断出现的老调重弹。比如,2012年,美国著名教授斯坦利·费什(Stanley

<sup>①</sup> Timothy Brennan, "The Digital - Humanities Bust", *The Chronicle Of Higher Education*, 2017 - 10 - 15, <https://www.chronicle.com/article/The - Digital - Humanities - Bust/241424>, 最近访问时间: 2018 - 5 - 27; 中文翻译参见顾佳蕙的《数字人文的幻灭》,“零壹 Lab (lingyilab)”, 2017 - 12 - 13。

Fish) 曾在《纽约时报》(The New York Times) 博客上发表名为《数字人文及其不朽》(The Digital Humanities and the Transcending of Mortality) 的文章<sup>①</sup>。他指出,数字人文所讨论的“作者”和“文本”不过是一种神学表述,提供的是一种关于新技术、新媒体可以改变印刷时代的线性思维困境的美好承诺。与布伦南观点类似,费什也认为数字人文不过是年轻学者试图在体制内部寻找更多可能性的政治性借口,其学术价值令人怀疑。

对于数字人文的批判同样来自数字学者阵营内部。比如,在信息文化研究方面享有盛誉的加州大学圣巴巴拉大学教授阿兰·刘(Alan Liu)就指出,在1960年代已经有大量关于信息技术革命的社会、经济、政治和文化批评,而当时的数字人文(“人文计算”)处于失语状态。刘强调,缺乏文化批评的数字人文是无法成为真正的人文研究的。数字人文必须超越“工具角色”,通过创造、改造和发展新的工具和方法,重建人文和公众之间的联系,才有可能对人文研究做出真正贡献。与布伦南对数字人文“技术/工具”持有否定态度不同,刘对于数字人文中的“数字”更为复杂。他认为,数字人文学者处于矛盾的心态之中,既担心数字人文被化约为工具后会降低其在方法论和认知论上的学术价值,又担心数字人文不够工具化,从而不符合其“创新”(innovation)和“建造”(building)的名号。

那么,数字人文是否仅仅在政治和技术上具有工具意义?数字人文学者自己怎么看待“数字/工具”问题?他们又如何看待自己的身份?

应该说,数字人文在过去二十多年间,经由早期的文本批评(textual critics),尤其是电子编辑(electronic editing)与人文计算(humanities computing)所开辟的路径,逐渐形成了一个具有丰富、多元且复杂的开放谱系格局。研究队伍的日益壮大,教学和科研机构的迅速增多,以及出版成果的井喷式出现,为其发展提供了机遇和条件,形成了一系列具有影响力的成果。与此同时,由于其内涵与外延的开放性,数字人文也一直处于不断被定义、被争论、被批判从而再被定义的过程,始终没有形成一个公认的明晰定义或特定的某些群体性特征;相反,“数字人文”更多的被认为是一种一定群体内的、具有共识性的、自我身份认知。而数字人文学者对数字人文中的“数字/工具”的看法,也随着数字技术的发展脉络和复杂的实践

<sup>①</sup> 参见斯坦利·费什《数字人文及其不朽》,王斌译,《文化研究》2013年第16辑,第199~205页。

路径经历了一系列变化,并不是一个简单的价值判断就能定论的。无论是布伦南,还是刘,都没有进入数字人文的“数字/工具”的内部去讨论其内在的合法性问题,更多的是将数字人文中的“数字/工具”黑箱化,变成了一个无须展开、毋庸讨论、不言自明的中间过程,而究竟这个实现是如何进行的,通过什么样的方式进行的,在他们的讨论中却没有讨论。比如,布伦南将数字人文简化为数字的“量化”,抹杀了数字人文中学科研究问题域的多样性。阿兰讨论工具化的时候,也没有具体说明工具化之所以发生的具体条件和内部过程。

然而,对数字人文中的“数字”进行讨论是非常有必要的。数字人文之所以不同于传统人文研究之处,正在于“数字”的背后代表的是一批学者试图以科学方法介入人文研究,从而建立新的认知方式、新的研究范式的自觉意识和实践。正如苏珊·霍基(Susan Hockey)在《人文计算的历史》(*History of Humanities Computing*)中所言,“人文计算不得不包含‘两种文化’,即将科学的严格、系统、明确、程序的方法特征带到人文学科中,来解决那些迄今为止大多以偶然的方式被处理的人文学科问题”<sup>①</sup>。然而,正如前面的争论所揭示的,这样的过程并非一帆风顺,水到渠成。一方面,数字与人文的结合并非仅仅是一种新的工具或者方法对人文研究中的施用:数字技术对人文研究的变化所起的作用是不可逆的,数字技术已经构成了目前人文研究的基础条件和环境,数字技术为信息社会中的知识生产转型提供了全新的维度和组织介入方式;另一方面,在过去几年,数字人文中对方法、工具的强调已逐渐将数字带向了更为模型化的“算法”模式。可以说,数字人文在试图以数字的工具、技术和媒介来实现学术话语的结构调整,改变已有的知识系统、生产与传播方式的同时,忽略了数字工具、技术和媒介本身所具有的差异性和意识形态,以及学术话语系统的社会语境,以及知识生产系统中的权力关系。

为了回应这种困境,本文采取从“数字/计算/工具”这一焦点问题为切入口,指出数字人文需警惕“算法驱动”依赖和“工具化”认识,需要突破在乐观和悲观意义上同时有效的“技术黑箱”,事实上,将“计算”问题视为单纯的算法、工具和方法,都无法解决数字人文的“计算”困境;

<sup>①</sup> 苏珊·霍基 《人文计算的历史》,葛剑钢译,《文化研究》2013年第16辑,第173~193页。

因此，数字人文需要从知识谱系上进行拓展，尤其需要重建包括媒介研究在内的其他思想谱系的关系，将数字人文放置到“数字文化”的大语境中，突破数字人文的工具化悖论。

## 数字人文：不可逆的数字知识生产路径

数字技术为信息社会中的知识生产转型提供了全新的维度和组织介入方式。如同印刷术之于西方文艺复兴、启蒙运动以及宗教改革等，基于印刷的书籍是信息革命之前最重要的知识存储、展示和传播的媒介和载体，信息与通信技术（Information and Communication Technologies，简称 ICTs）也成为 20 世纪中期以来的“网络社会”（Network Society）中的基础性因素，推动着知识生产系统从基于印刷向基于数字媒介（digital media）的转变。在曼纽·卡斯特（Manuel Castells）所言的网络社会中，新的信息模式改变了传统社会的时空关系，压缩了时空距离，而数字技术的瞬时性使得在不同地点和不同时间的人们可以同时处于一个信息网络中。与此同时，信息从其具体化的对象中被抽象出来，得以在不同环境、媒介中自由流动，而基于信息与通信技术的数字媒介也成为最主要也最重要的信息传输、存储和展示方式，改变了人们获取、理解和生产信息的方式，改变了包括学校、图书馆、博物馆等在内的文化、教育机构及学术共同体的生产、组织和传播知识的方式。

借用海尔斯（N. Katherine Hayles）的观点，这个过程可以总结为一种处于从基于印刷（print-based）向基于数字（digital-based）的知识生产转型期。在这个转型过程中，知识生产方式发生了变化。随着个人计算机的普及、互联网的发展、搜索引擎和数据库的迭代升级，无论是在日常生活还是在工作中，越来越多人文学者适应并习惯了每日通过谷歌（Google）、必应（Bing）、百度这样的搜索引擎，或在 JSTOR、Project Muse、CNKI 这样的学术数据库中查找资料，为研究进行基础性的准备。与此同时，慕课（Massive Open Online Courses，简称 MOOC）等在线开放式教学方式也进入了国内外高校的教学体系中，吸引了学生与教师们的大量参与。再者，搜索引擎及其背后的算法参与了学术生产的过程。在国外，谷歌学术（Google Scholar）已经成为重要的学术信息搜索引擎与平台，而脸书（Facebook）、Academia.edu 等类社交网站也逐渐取代了过去的邮件列表服务（mailing list

services), 成为新的学术社群聚集地; 在国内, 百度、必应等搜索引擎, 虽然在学术搜索方面表现不佳, 但像中国知网 (CNKI) 这样的学术数据库, 也已经开始提供“文献分析”及可视化效果, 为学者提供一定的数据分析。此外, 豆瓣和微信公众号等社交网络平台, 开始深入学术社群营造及学术传播过程中, 在译介、传播、评价、讨论等方面发挥着越来越重要的作用。

人文学者在这个过程中的参与程度不同, 有的人仅仅使用数字媒介来进行邮件交流、网络搜索或使用电子资源进行写作; 而有的人则更多使用网络搜索和在线网站来获取学术资源, 使用数字技术作为研究过程中的一部分, 比如, 发表论文到网上以期扩大读者群和影响范围; 部分数字人文学者从一开始就使用数字媒介来概念化 (conceptualizing) 和实施 (implementing) 研究项目。但无论是何种程度, 基于印刷的学者, 或是基于数字的学者, 都已经卷入了这一场数字知识生产的变革中。更根本性的是, 有研究证明, 与网络的互动, 哪怕是点击鼠标这样的小动作都会对人的神经系统产生影响。我们对数字介入得越深, 我们对具有联网和编程功能的机器所具有的执行复杂认知能力就了解的越多。学者研究项目在研究和教学中的介入也对物理的教学和研究空间、方式提出了新的要求, 因此, “在这一点上, 很难说这场转型究竟在哪里结束, 几乎人文研究工作的方方面面都需要被考虑, 这包括了研究和出版、教学和指导、评价和同行评议以及学术界与更大的社会之间的关系等”<sup>①</sup>。数字人文可以说是这样的一个学术生产的社会语境及场域中出现的具有一定必然性的选择。

综观数字人文的历史, 从布萨神父 (Roberto Busa) 与 IBM 在 1949 年的合作开始, 到 1980 ~ 1990 年代的语料库语言学、计算语言学再到 1990 年代中期之后的人文数据库、档案库和量化分析研究等, 数字技术与人文研究的弥合、交融已经形成了一个清晰的发展脉络和路径。<sup>②</sup> 自 1949 年以来, 数字工具、方法与认知方式就在不断地改变着人文研究的走向, 在传统人文研究内部和外部都开拓出了新的路径: 无论是倡导以数据库重构基于印刷文本和电子文本互文性、内构性的文本批评 (text critics), 或是通过数字化、数据化建立一种量化的、可计算的、基于模型建构的“人文计算” (humanities computing), 数字人文内部已经形成了尽管宽泛但依然有迹可循

① N. Katherine Hayles, *How we think: digital media and contemporary technologies is*, Duke University Press, 2012. p. 5.

② 参见苏珊·霍基《人文计算的历史》, 《文化研究》(第16辑), 2013年9月, 第173~193页。

的、有理可依的理论与方法，也积累了一大批在理论和研究方法上均具有一定原创性的学术成果。例如在英语世界中，就有以弗吉尼亚大学麦根（Jerome McGann）的“理性超文本”（Rationale of Hypertext）开创的“文本批评”、以芝加哥大学莫莱蒂（Franco Moretti）及其在2008~2013年完善形成的“远读”（*Distant Reading*）模式、米歇尔（Jean-Baptiste Michel）团队基于谷歌图书的“文化测量”（*culturomics*）模式、以列维·曼诺维奇（Lev Manovich）和马克西米利安·席希（Maximilian Schich）为代表的“文化分析学”（*Cultural Analytics*），等等。这些研究项目无论是在方法论革新性或者是研究成果原创性上都可谓典范。以莫莱蒂的“远读”模式为例。由于其具有的“显而易见”的针对性（与西方传统人文研究近百年的“细读”的区别）、行之有效的示例解释性（达到了传统研究的标准，并有新的“发现”）以及超尺度的分析范围（大尺度的文本量）使得“远读”成为数字人文领域中最具传播力并最具认可度的概念之一，也被公认为在理论和实践意义上都有一定原创性的数字人文的“范式”。<sup>①</sup>其后又有苏真（Richard Jean So）和霍伊特·朗（Hoyt Long）在“远读”概念的基础上提出关于日本俳句的“文学模式识别”（*Literary Pattern Recognition*）研究，在检验一个已知结果的正确性之外，还以一种新的计算方式挑战及改变以往对于俳句的认知及研究思考。莫莱蒂及其拥趸所进行的相关研究已经难以用原有的文学研究的模式来讨论，更重要的是，他们依据新的研究方法不仅实现了传统人文研究的结论，而且发现并回答了以证据、经验和阐释为主的传统人文研究所无法发现或回答的问题。

汉语世界的数字人文尽管起步略晚，但在过去的十多年取得了长足的进展，其中执牛耳者就有哈佛大学包弼德团队基于“中国历代人物传记数据库”（*China Biographical Database Project*，简称“CBDB”）开展的“群体传记学”（*Prosopography*）、以台湾大学项洁（Jieh Hsiang）团队开展的以数据库带动文本间脉络的大图景研究、以曾在香港中文大学和台湾政治大学任教的金观涛、刘青峰带领开展的“观念史图像中的事件”研究等。尽管

<sup>①</sup> 参见美国现代语言协会的会刊 PMLA 2017 年第 2 期十位学者对莫莱蒂的《远读》一书的专题讨论。特别是在安德鲁·哥德斯通（Andrew Goldstone）的《阅读的常识》（*The Doxa of Reading*）一文讨论到了在 2008 年莫莱蒂发表了最早关于“远读”的文章后，该概念是如何被接受和传播的。另参见本专辑但汉松的文章《朝向“数字人文”的文学批评实践：进路与反思》。

发展路径不同,方法各异,但这些研究也针对中文语言和文本的特殊性,形成了一系列具有各自特点、路径和模式的研究成果。比如项洁团队以唐宋两部官修大型类书《艺文类聚》和《太平预览》为例建立起的全文资料库及其查询、分析和观察功能为研究者提供了从一个宏观视野观察唐宋士人知识结构中渐进而深刻的变化轨迹。这种研究提供的对历史文献的分类结构、比重和条目内容上的差异的观察和分析,非人力能及,恰恰是通过数据库系统和相关的后台计算功能才得以实现的。<sup>①</sup> 无论何种语言,我们可以看到数字人文学者对于这种以科学方法介入人文研究从而建立新的认知方式、新的研究范式有着非常清晰的自觉认识。莫莱蒂研究中就强调量化数据对于理论的证伪性意义。而金观涛则提出,数据库和数字人文学使得人文研究成为可检验的,而且这种可检验性与社会科学中广泛使用的量化分析不同。而更为激进的斯蒂夫·拉姆齐(Stephen Ramsay)则高举起了“算法批评”(Algorithm criticism)的大旗,呼吁一场变革,这场变革关心的并非仅仅是方法或者程序上的改变,更是希望能寻找到对于文本研究特定研究效果更为关注的读者。<sup>②</sup>

由此,我们可以看到数字人文不仅是信息社会的知识生产的可能性之一,也同时成为知识生产转型的驱动方式之一。这是信息技术、数字技术在当代社会和发展中所带来的知识生产的必然呈现,无论是不是被叫作“数字人文”,我们都必须要面对这样一个数字知识生产的转型。与此同时,也可以看到,从宽泛意义上的数字、工具、量化方法到更为具体的“算法批评”,数字人文中“数字”的含义变得越来越依赖于算法,这既带来了机遇也带来了风险。

## “算法”困境

随着人文计算的被认可度远超过校勘学<sup>③</sup>,作为核心驱动力之一的计算

① 参见项洁、陈丽华、杜协昌、钟嘉轩《数位人文视野下的知识分类研究——两部官修类书的比较分析》,《东亚观念史集刊》2015年第9期,第229~286页。

② Stephen Ramsay, *Reading machines: Toward an Algorithmic Criticism*, the University of Illinois Press, 2011, p 17.

③ 校勘学者甚至提出两者应该划清界线。参见Robinson, Peter, “Why Digital Humanists Should Get Out of Textual Scholarship. And If They Don't, Why We Textual Scholars Should Throw Them Out.” *Scholarly Digital Editions* 29 Jul. 2013.

(computation) / 算法 (algorithm) 在数字人文研究中就变得越来越重要。<sup>①</sup> 不过, 也出现了诸多争议、论争和潜在的危机。因此, 我们需要理解在数字人文研究中, “计算” 代表着什么。在霍基所论述的人文计算的历史中, 这个概念涉及的不仅包括了语言分析、数据标准、编程语言和数据库等方法, 还包括了软件的操作系统和计算机等工具和硬件研发。但在近二十年来, 数字人文语境中的“计算”面貌已经发生了变化。

首先, 霍基所谈到的早期人文计算中语料索引、语料库和文本编码等逐渐成为一种公共基础性的、基于机构平台的工作, 与研究者个人渐行渐远。甚至当年霍基所高度看好的 TEI, 也因为其烦琐性、复杂性和高成本已经不再受到数字人文学者们的偏爱; 其次, 霍基所认为的计算语言学与人文计算的分离, 却并没有那么彻底, 计算语言学在数字人文的知识结构中依然占有重要的地位, 尤其是语料库语言学已经发展成为非常成熟的学科<sup>②</sup>; 最后, “计算”更多地成为一种可以抽离并施用于不同文本的应用方法和模型, 或者说, 算法。不管是自然语言处理 (Natural Language Processing), 还是统计分析, 不管是语义分析, 还是情感分析; 不管是文本挖掘, 还是图像处理, 研究者在通过各种计算方法重复人文学者的研究路径的同时, 试图建造非计算不行的新型研究模式。这种尝试不乏很多成功案例。比如, 以米歇尔为首的研究团队与谷歌图书合作开展的“基于百万数字图书的文化量化分析”就是基于 5195769 本数字化书籍的语料库的计算分析 (computational analysis), 以量化的方式呈现文化趋势。研究团队以自然语言处理中较常用的 n-gram 模型方法, 以单个词或多个词为单位, 来自全世界的大学图书馆的 1500 万本数字化图书中选择了从 1800 年到 2000 年的 500 万本, 共计 7 种语言 500 亿字的文本量进行了统计分析, 对包括英语词汇量变化、英语语法的变迁、集体记忆与健忘、大众声望、审查检测等文化议题进行解读。因为是基于两百年间的词频波动观察, 所以得出的一些结果还是非常具有启发性的。<sup>③</sup> 这种通过对大数据集进行定量分析从而学习人类

① 核心驱动力应该还包括“数据 (data)”“数字化 (digitization)”和“可视化 (visualization)”等, 但因篇幅有限, 且已有诸多论述, 此处不再涉及。

② 参见金雯、李绳《“大数据”分析与文学研究》, 《中国图书评论》2014 年第 4 期。

③ 比如英语书籍中最常使用的单词量实际上比权威字典要多, 而且其中大约 63% 的英文单词在齐夫定律 (Zipf's law) 的测量下是低频使用词, 更有 52% 是没有被收录到词典中的。

文化的方式被命名为“文化测量”模式<sup>①</sup>，并于2011年在《科学》杂志发表。此后，不少学者也将此模式施用于了不同的文化数据集。<sup>②</sup>比如卡莱弗·李塔鲁 (Kalev H. Leetaru) 对过去三十年全球的本地新闻进行了调性和地理分析，并成功预测了2011年在阿拉伯半岛发生的重大政治事件以及该事件发生的地点。<sup>③</sup>另一种从宏观视角观察文化历史的模式是由美国得克萨斯州大学达拉斯分校的马克西米利安·席希 (Maximilian Schich) 所带领开展的两百年间世界 (主要是欧洲和北美大陆) 的文化历史“元叙事” (meta-narrative) 问题。<sup>④</sup>

这些研究优势显而易见：首先，以一种模型化的方式从宏观角度研究人力所不能及的文化问题。但我们也要意识到数字人文将“算法”抽离于具体语境、对象和文本之外建立数学模型的危险。比如，文化测量模式仅在谷歌图书语料库这一特定对象上有意义，而在其他文本数据上并无效力。其次，对研究模型和方法的强调，使数字化和数字处理过程中大量烦琐的基础工作一笔带过。比如，数字化过程中涉及文本物理属性的数字化标准、数据文本的OCR识别错误、数据库的兼容性和数据格式的混乱、数据清理过程中的人力投入，等等。这些都被视为基础准备的前期工作，而不被认可为有价值的研究部分。然而，数据结果呈现效果往往取决于这部分的工作。再次，对计算或算法强调，一定程度上造成了数字人文研究中的“技术黑箱”。在控制论中，“黑箱理论”用以指称抽象再现一个开放系统的方式，这个系统中的刺激输入和输出反应是可以观察到的，但这个“盒子”

① Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. Quantitative analysis of culture using millions of digitized books. *Science*, Vol. 331, 6014, 2011, pp. 176-182.

② 2011年之后，多位学者用谷歌数据集和Ngram Viewer进行了基于英语及其他语言的语汇分析。参见 Alexander M. Petersen, Joel Tenenbaum, Shlomo Havlin & H. Eugene Stanley, “Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death,” *Scientific Reports*, Vol. 2, 2012; Roth, S., Fashionable Functions. “A Google Ngram View of Trends in Functional Differentiation (1800-2000),” *International Journal of Technology and Human Interaction*, Vol. 10 No. 2, 2014.

③ Leetaru, Kalev H. “Culturomics 2.0: Forecasting Large-Scale Human Behavior Using Global News Media Tone In Time And Space,” *First Monday*. 16 (9).

④ Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, Dirk Helbing, “A network framework of cultural history,” *Science*, Vol. 345, Issue 6196, 2014, pp. 558-562.

系统的结构与被观察到的输入、输出路径并不具有相关性。<sup>①</sup> 在计算机领域中，“黑箱理论”也被广泛使用，尤其是以隐喻性的方式指称在高度模块化的封装系统中，用户能观察到的只是输入的信息和输出的结果，而其具体计算的过程（物理层面或是源代码层面，甚至编程语言的层面）都是无法被用户完全认知到（认知的程度取决于用户的身份，终端用户还是不同层次的程序员）。正是在这个意义上，数字人文中确实存在布伦南和刘所批评的“工具主义”或者“算法过度依赖”的问题，将算法部分“技术黑箱”化，没有真正深入到“工具”或者“算法”的内部去考察、展开，由此做出的判断，也就不能令人信服。

## 媒介认识论突围

如何突破数字人文的“算法困境”？一种是从数字人文的内部，强调研究过程、技术方法的透明化、可验证性和语境性，提倡元数据标准化、数据库开放、数据标准兼容、数据处理方法可重复性等，更不能因为强调算法而忽略人工介入的大量工作。尤其是在知识生产转型阶段，数字知识生产方式尚未全面建立，数字基础设施尚不完善，数字知识共享机制非常缺乏的情况下，过度强调算法实际上对数字人文实际上有害无益。正如金观涛和刘青峰所明确提出的，“数字人文”是需要人工的文本语境与数字方法同时进行的，而这一点也在一定程度上说明了当下的数字人文的人文研究属性。

另一种困境突围的可能性存在于数字人文的外部，即通过拓展数字人文的思想谱系，接纳包括媒介研究、图像研究、游戏研究等在内的思想及方法，从而构建一个以“数字文化”为基础的大数字人文观念。以媒介研究为例，其将从根本上颠覆数字人文争论中的“工具”问题。

从媒介研究认识论的角度重新构建数字人文中的数字与计算问题，则有另外一种路径。计算机作为一种媒介参与数字知识生产，其提供了结构的数据化基础和界面隐喻的视觉呈现，重构了知识生产的底层设计方式。与书籍不同，计算机作为媒介，其存储和展示功能是分离的。由于计算机

---

<sup>①</sup> Mario Bunge, “A general black – box theory”, *Philosophy of Science*, Vol. 30, No. 4, 1963, pp. 346 – 358.

的信息处理方式可以将几乎一切事物都可以以0和1的二进制代码进行数字化,因此,各种类型的文字、图片、声音等都可以作为数据被储存起来。在文本生成层面,数据化结构成为数字书写的深层生成结构,同时也决定了我们通过什么样的结构来描述我们需要描述的现实事物。深层的数字书写则成为数据库的设计。计算机界面则提供了一种将这种数据组织结构的隐喻性呈现方式。在基于印刷的文本中,后台数据库和前端的界面是一致的,或者说界面是不存在的,我们直接面对的就是被组织起来的数据,也就是由文字构成的叙事。而在数字文本中,界面则成为深层数据组织结构的直观呈现。比如,叙事的“用户”正是跟随由数据库创造者建立的数据库记录之间的链接来穿越数据库。而一个交互叙事就可以理解为通过一个数据库的多重轨迹的总和。而一个传统的线性叙事就是多个可能的轨迹中的一个。

媒介研究将有助于重构“算法”问题。从媒介研究的角度来看,算法并非是计算机或者数字对人的操纵与愚弄。综观计算机语言发展历史,可以看到,计算机语言则经历了从计算到逻辑再到编程语言的一个过程,与人类的思维共享了一种认知框架。莱布尼兹所设想的理想的“通用符号集 (Characteristica Universalis)”能够将有意义的陈述或者推理片断翻译成一种逻辑演算,从而将人类的各种语言带入单一的共享数据库。这种适合数学计算的计算机原型尽管不能用他设想的二进制计算逻辑进行演绎证明,但已经是一种模仿了上帝思考的前现代的人类智力模式,奠定了现代逻辑作为符号科学的基础。<sup>①</sup>自从图灵机的发明到冯·诺依曼数字计算机的问世,计算机从对人类认知方式的模拟走向了数字化。当布尔逻辑被用于计算机的搜索和查询时,从语言到逻辑的“翻译”过程被计算机所取代了。“人的思维也用不着再去想语言如何才能适合现代逻辑的体系,计算机把这些任务承包了;它将我们的字母变成了机器可操作的数码字。”<sup>②</sup>计算机成为我们获取信息的窗口和漏斗,我们在对计算机提问时,要采取适应于信息世界提问的方式,比如,我们在计算机上搜索时,必须要选择适当的“关键词”,关键词选择的不同直接影响到我们获得信息的数量、相关性和有效性。这促使我们学会用计算机的语言说话,要去适应一种新的“知”的方

① 海姆《从界面到网络空间——虚拟实在的形而上学》,上海科技教育出版社,2000,第35~36页。

② 同上书,第19页。

式。在1957年，当计算机还处在只能进行数值运算的UNIVAC阶段的时候，海德格尔从哲学的角度在对技术的本质进行反思的时候，就注意到了语言机器是现代技术对于语言模式和语言世界进行控制的方法之一，而并不像人们想象的，人是语言机器的主人，而可能是恰恰相反，语言机器控制了人类的本质。这主要是因为计算机正在对我们的思想过程进行着改造。计算机技术不仅灵活，而且容易适应我们的思想过程，所以我们很快就不再把它当成一种外部工具，而是更倾向于把它视为第二皮肤或精神假体。一旦适应了技术，我们便会像音乐家玩乐器似的玩起技术来，认它为同一，与它合二为一。在语言机器上进行写作，产生出一种新的写作和思考的方法。我们的指尖所及便是帕斯卡和莱布尼兹——现代形而上学之父——梦寐以求的一台演算机器，但现在这台机器却是对我们的语言进行操作。

尤其到了编程语言（Programming language）出现以后，无论是在二进制代码中，还是在界面的生成过程中，代码（code）都或显或隐的存在并且发挥着作用。代码本身具有书写的特征，需要运用一定的语言规则，因此编写代码也就成为一种书写形式，是一种导向某种结果的工具。在代码编写的过程中体现着操作者与机器之间的互动，而意义就在参与代码的写作过程中被凸显出来，在编码中，甚至有时候我们将错误也视为创造意义的一种方式。就像海尔斯所言，“编码错误实际上指向的是系统内的不可避免的噪音，与符号的概念化，即作为符号不同层面之间的一系列转型的产品相联系的转移”<sup>①</sup>。这实际上指向了从数据库到代码的过程，代码作为一种可执行的语言，也是介于媒介物质性和编程语言之间的中介。一方面，代码的物质性体现在它是由磁性媒介上的记号组成，记号则是通过像键盘、鼠标、触摸屏和其他设备被铭刻在磁性媒介上；另一方面，代码作为一种语言，是再现的。这些记号的序列通过语言、句法和表达逻辑来传达思想。

由此对数字人文进行反思，我们就会意识到，数字人文的知识生产并不仅仅在文本的层面进行，更重要是在数据库、代码和底层计算的层面进行。而人文学者在一定意义上是与数据库语言、编程语言乃至物理电路进行合作开展研究。数字人文的“实践性”并不仅仅在于研究中如何处理词频，或者如何使用分析软件，而恰恰在于，在参与的过程中，尤其是在在更

<sup>①</sup> N. Katherine Hayles, *My Mother Was a Computer: Digital Subjects and Literary Texts*, Chicago: University of Chicago Press, 2005, p. 68.

高层次的参与过程中,数字人文学者是通过计算机语言,从开始就参与了他所要设计的数据库、数据方法和数据分析路径的意义解释与建构过程。研究者必须将自己的想法转换成机器可以理解的命令,并对机器做出的反馈进行下一步的判断和设计。而机器在这个循环中则扮演了一个模拟认知的角色,它通过对编程者指令的执行和回应,从而在计算机虚拟空间中模拟出编程者的想法,并以同样的语言和表达方式呈现出来。与此同时,这样一个反馈循环的过程赋予了其最大程度的弹性。“这主要体现在最初的设计、不可避免的修改、不断进行修正以及维护大系统的需求。(程序语言中的‘动词’因此成为过程,对象通过它能够在相互之间以及和系统设计之间互动”<sup>①</sup>。因此,人文研究的计算过程也具有了很大可能的机动性,换句话说,我们可以介入代码的编写和高级语言之中,通过修改代码或者输入新的指令来改变系统从而改变数据处理的方式,这个过程本身就是一种人与机器的共同互动,人与机器在此分享的是同一种语言和认知。

至此,我们回到一开始布伦南对数字人文的质疑,我们可以非常明确地回应,数字人文学者,包括所有使用计算机或者计算终端的学者们,在使用计算机或者计算设备在进行书写、创作和研究的时候,就已经从大脑的层面建立了与计算装备的一个反馈回路。不管我们是否真正“建造”(building),我们都已经是数字人文学者。

## 小 结

本文选择有关数字人文的系列争论为横切面来介入数字人文的交锋处,力图在“深描”之中揭示思想交锋,彰显“数字人文”领域的多样性和复杂性,并通过媒介研究来反思数字人文知识谱系扩展的可能性。然而,数字人文中的计算问题实际上比文中谈及的要复杂得多,比如如何看待“量化”与数字人文的关系问题。在“算法”主导的诸多数字人文研究中,算法与量化基本是被同等视之的。这就将数字人文进一步带入了一种抽离了具体社会、文化甚至技术语境的自为境地,而对于整个社会知识型的转型和建设而言,并无益处。在这个意义上,《数字人文宣言2.0》(*The Digital Humanities Manifesto 2.0*)的观点则更具有启发性。在这篇由杰佛瑞·施纳

<sup>①</sup> N. Katherine Hayles, *My Mother Was a Computer: Digital Subjects and Literary Texts*, p. 75.

普 (Jeffrey Schnapp) 和托德·普莱斯勒 (Todd Presner) 牵头、彼得·路勒福尔德 (Peter Lunenfeld) 与乔汉娜·朱可 (Johanna Drucker) 及数字人文众人参与的“宣言”中, 数字人文被认为具有一种聚合实践的普遍性特点:

印刷不再是一种知识生产及 (或) 转播的唯一或标准媒介; 相反, 印刷发现自身被吸收进了新的、多媒体形态之中; 而数字的工具、技术、媒介则深刻地改变着艺术、人文和社会科学中的知识生产与传播。数字人文寻求在世界中发挥首要作用。在这个世界中, 大学, 而不再是知识和文化的独家生产者、管理者或者传播者, 被号召起来去为当下新兴的公共领域中的学术话语打造天然的数字模式, 去树立这些领域中的卓越和创新, 去实现全球和地方的即时知识生产、交换和传播网络形式。<sup>①</sup>

---

<sup>①</sup> “The Digital Humanities Manifesto 2.0”, *Digital Manifesto Archive*, <https://www.digitalmanifesto.net/manifestos/17/>, 最近访问时间: 2009年6月22日。