

影响力报告： 发现 Gale-ASECS 奖学金怎样帮助 研究者利用数据 做出新发现

阅读其中一位奖学金获得者怎样使用数字人文工具探讨十八世纪英国的决斗历史



Adam Kozaczka 博士

美国德克萨斯农工国际大学英语系助理教授，
Gale-ASECS 奖学金学者

认识学者

作为 2022 年获得 Gale 和美国十八世纪研究学会（ASECS）奖学金的五位研究者之一，Adam Kozaczka 使用了数字人文（DH）工具提升他对十八世纪英国决斗文化的研究。通过探讨来自相关法律议题和文献的数据，Kozaczka 想要更准确的描绘出决斗在英国的流行以及十八世纪有关决斗的评论。

所有的 Gale-ASECS 奖学金获得者都获得了 Gale “十八世纪作品在线”（Eighteenth Century Collections Online, ECCO）和 Gale 数字学术实验室（Gale Digital Scholar Lab, 简称“实验室”）六个月的使用权，用以开展研究。ECCO 是最具综合性的在线历史档案，支持十八世纪研究，让研究者能够访问到 1701 至 1800 年之间英国印刷出版的每一部重要英文和外文图书。ECCO 中囊括的海量内容与消除了数字学术入门障碍、业界领先的文本和数据挖掘平台 Gale 数字学术实验室的巨大研究潜力结合在了一起。

学者们经过培训，掌握了实验室中提供的文本和数据挖掘工具的使用方法，以数字人文的研究方法探讨 ECCO 并推动他们的研究工作。阅读本文，了解 Kozaczka 是怎样利用这个基于云计算的分析平台来构建可靠的数据集，并使用数字人文工具运算得到有意义的分析结果，从而做出新发现的。

找到更多历史背景资料

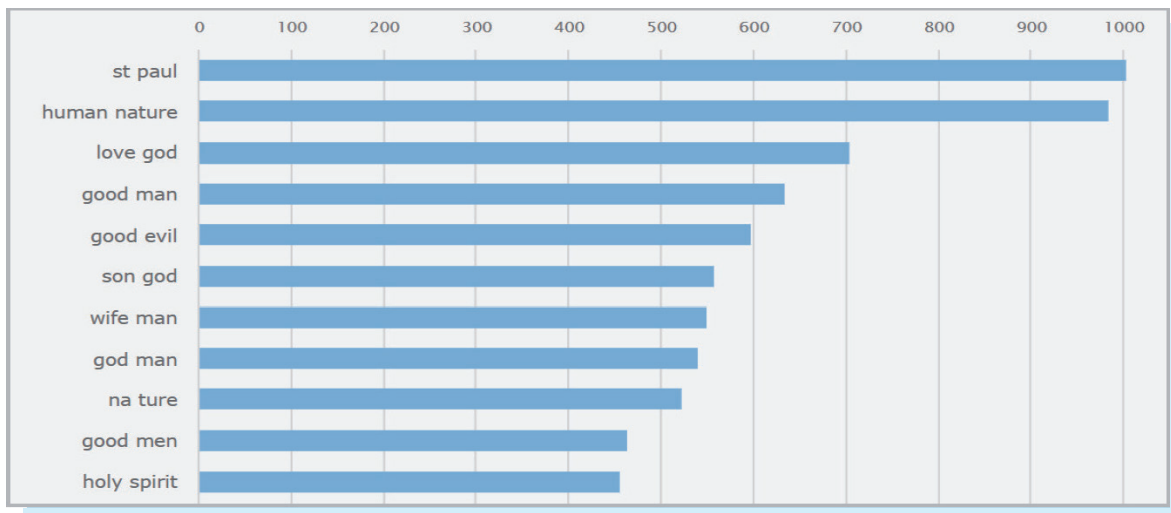
为了清晰了解十八世纪英国的决斗历史，Kozaczka 在作为 Gale-ASECS 奖学金学者期间，通过培训并利用他的访问权限，整理和分析了更完整的一套那个时期对决斗的评论观点。

根据 Kozaczka，“这个项目的目标是使用实验室中提供的工具筛选 ECCO 中与十八世纪的决斗相关的大量文献，检验、拓展和强化历史学家、文学评论家和法学学者得出的结论，这些学者都曾在没有数字人文工具的帮助下完成了关于决斗的重要研究著作”。

Kozaczka 的研究包含两个基本的问题：“决斗是否在英国很普遍，或是在某些城市或地区决斗这种行为发生的可能性更大？哪些作者曾书面论述过这种行为，他们使用了哪些修辞方式？”

Gale-ASECS 奖学金为 Kozaczka 提供了机会，发现更丰富的与其研究相关的数据，并可以使用强大的数字人文工具分析数据。

“实验室激发的游戏感和实验感促进了对 ECCO 的深入探索。”



“决斗（论述）” [Dueling [polemics]] 内容集中的二元语法 (Biograms)。

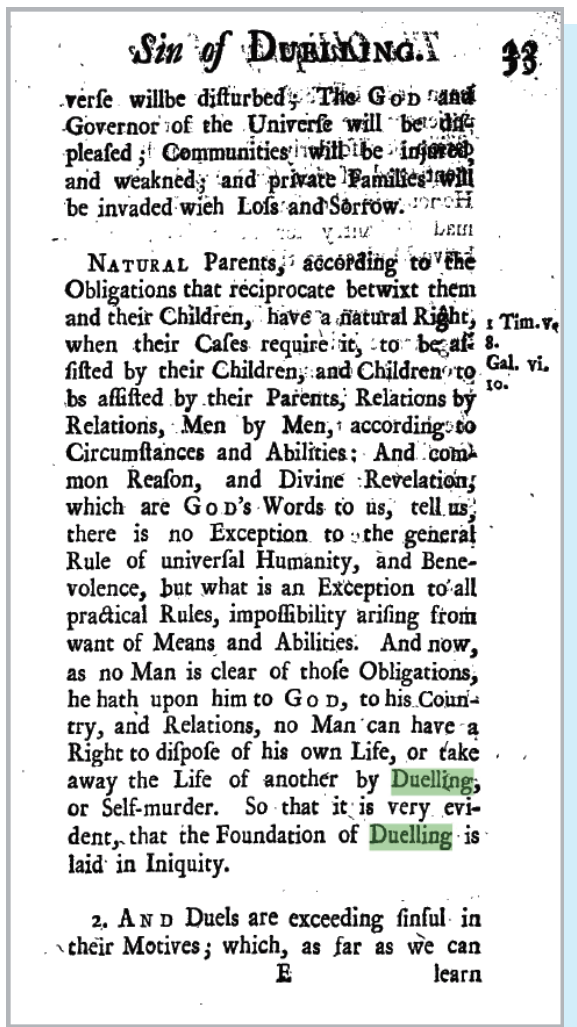
利用数据做出新的发现

在他的研究中，Kozaczka 想要解决一直存在的、围绕十八世纪英国决斗历史的分歧。他在他的报告中详述了这种分歧，“社会历史学家 Donna Andrew 在 20 世纪 80 年代到 21 世纪初完成的有关十八世纪决斗的著作描述了反决斗活动家和决斗支持者之间在印刷文献上持续不断且势均力敌的辩论……Andrew 引用了很多赞同决斗的资料，指出这是一场两方面的斗争，而有时她则夸大了赞成决斗的一面。”

然而，Kozaczka 在实验室中开展研究时收集的数据则让他有了一项新的发现。“利用实验室工具筛选 ECCO 中的文献，呈现出完全不同的结果，”他说。

“首先，宗教至关重要：在我名为‘决斗（论述）’的内容集中包含的文献阐述了对决斗看法（不同于有关决斗的法律专论或涉及真实决斗或决斗审判的事实报告），其中最常出现的一元语法是‘god’（上帝），出现了超过 32,000 次，”Kozaczka 指出。“最常出现的二元语法是‘St Paul’（圣保罗），出现 1,000 多次，出现次数最多的前五个二元语法还包括‘love god’（爱上帝）、‘human nature’（人性）和‘good man’（好人），显示出表面上论述决斗的文献实际上在宗教和道德语言上花费了比法律术语更多的时间。”





论述不洁和决斗之不道德的基督教文章，在圣保罗大教堂的两次布道中宣读。艾塞克斯郡沃尔瑟姆教区牧师 Anthony Holbrook 著，1727 年，ECCO。

挖掘被忽视的故事

进一步分析这些数据，Kozaczka 发现，“当我根据这些一元语法和二元语法返回到文献原文时，我发现几乎所有这写文献都是反决斗的。而且，我能够追溯到的那些作者都是持异议的大臣们，或至少曾对其他往往与宗教议题无关的问题发表过短评。因此我们开始看到在更大范围内、更多议题上对社会正义的追求，在这种情况下，谴责决斗行为的人也会对其他议题发表见解——例如奴隶制，如我们所知，英国的废奴运动很大程度上是由新教争议中持异议的一方推动的。”

Kozaczka 在实验室中对来自 ECCO 中的大量文献进行研究后发现，“我的‘决斗（论述）’内容集收集的文献包含这些争论，但几乎没有几部作品对决斗这种行为有积极的表述。这揭示出 Andrew 的著作虽然逐期翻阅了数十年十八世纪期刊的微缩胶片（她这样描述了这个过程，可能并不能检索关键词），但仅使用了来自辩论双方的部分代表性资料。”

Kozaczka 总结说，“我的内容集覆盖范围更广，意味着我能够立刻看到两方之间的差异，这种差异直到今天才被学术界认可。”

“使用实验室的机会令我能够更加深入的探索 ECCO 的资源，比我之前认为的还要更深入。”

利用数字人文方法拓展研究可能性

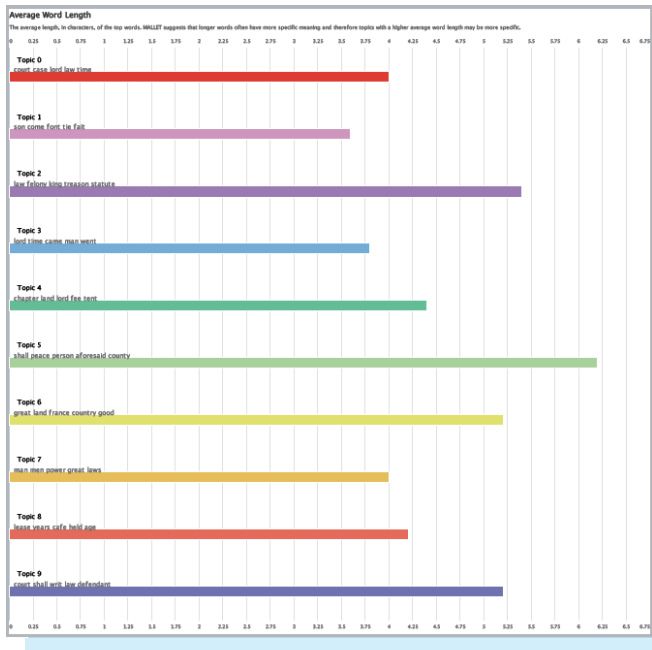
在整理他的数据并做出初始的发现后，Kozaczka 指出，“实验室的分析功能让我能够从这个内容集中获得更多信息。”

“实验室带来的益处无论怎样强调都不为过，六种分析工具让不具备硬技能（例如编程）的学者能够轻松使用数字人文的研究方法，而正是这些硬技能曾经将数字人文学者与——可能没有更好的称呼、我称之为——一般学者或传统学者区分开来，”他说。

Kozaczka 的经历说明，像实验室这样的数字人文工具取得的长足进步为新手和经验丰富的研究者们消除了通向数字学术的障碍，让他们能够发现和分析与他们的研究更紧密相关的数据，研究结果更有意义。

“实验室是下一代数字人文平台的早期范例，它将数字人文研究方法加入到每位学者的工具箱中。”





“过失杀人”（Chance Medley）内容集——主题建模结果

利用分析工具发现新见解

在详细阐述他用来提升研究的数字人文工具时，Kozaczka 主要围绕 N 元语法（ngrams）和主题建模（topic modeling）工具。

“每一种工具有多大的用处很大程度上取决于正在探讨的内容集，”他指出。“我的内容集‘决斗（论述）’、‘决斗（期刊）’和‘决斗（创意写作）’，利用 N 元语法工具得到了很多有用的结论。

“而另一方面，‘过失杀人’内容集则生成了主题较为明确的主题建模结果，因为‘过失杀人’是一个很明确的法律词语，”Kozaczka 举例说明。“因此这个词语常常出现在审判报告中，最常出现在法律专论中（不同于‘决斗’或‘决斗者’，这两个词在各种文献中均会出现），意味着在‘过失杀人’内容集中的文献有着共同的词表，而在‘决斗’内容集中的文献则不然。‘过失杀人’内容集应用主题建模工具后得到了极为有帮助的主题，例如叛国、案件、法令、重罪、国王等，帮助我找到具体的政治审判和报告。”

为未来的研究和教学奠定基础

Kozaczka 在 Gale-ASECS 奖学金期间开展的研究让他构建起一个范围广泛的数据集，支持他未来的课堂教学和对十八世纪决斗的进一步研究。

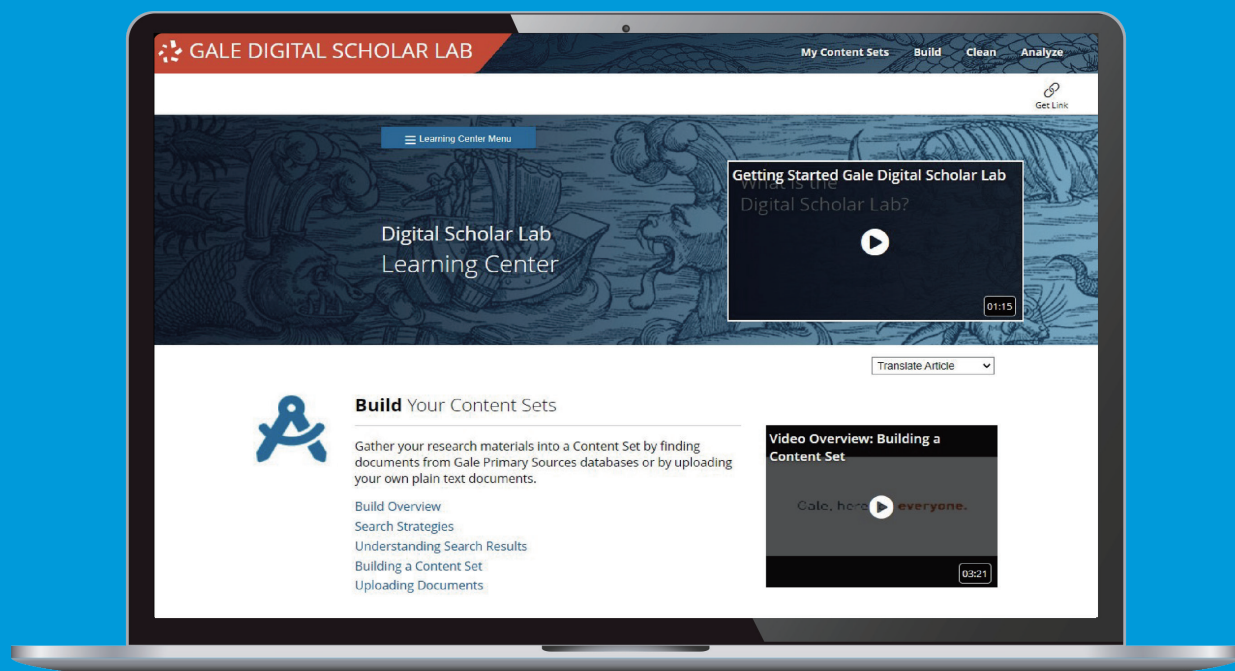
“我有很多研究项目结果的屏幕截图，我将会找到一种方式在讲课时展示给我的学生们，”Kozaczka 说。“我已经发现，数字人文对于有时（偏颇地）认为人文学科是大学教育‘较软弱’一面的二十一世纪学生而言非常激动人心。认真对待技术但对着莎士比亚打哈欠的学生们能够在数字人文研究项目中变得异常兴奋，我认为实验室既具有象征性，也易于展示，足以吸引他们的注意力。”

Kozaczka 指出，“我将使用我在奖学金期间收集的一些信息进一步探索数字人文。

“我已经下载了我内容集中大部分资料的 OCR（光学字符识别）文本，我的目的是将所有这些文本放在一起，对它们运行 Python 程序，尝试进一步分离出更多信息，”Kozaczka 说，勾勒出他的下一步研究。“也许更为重要的是，我在这一研究项目中已经得到的很多结果将会成为我正在写作的一部法律和文学专著的一个章节，”他总结道。

“在这项奖学金期间完成的工作将继续影响我未来一段时间内的教学方法和学术研究方向。”





查看更多有关 Gale 强大研究工具的详情

想要对 Kozaczka 用来提升研究水平并培养数字人文技能的 Gale 资源有更多的了解？

发现 Gale 数字学术实验室带来的各种可能性：

gale.com/digital-scholar-lab

找到更多有关 ECCO 的信息：

gale.com/ECCO

想要更多了解美国十八世纪研究学会？

AMERICAN
SOCIETY FOR
18TH-CENTURY
STUDIES

发现更多有关 ASECS 的详情，包括学会的项目计划和活动等，请访问：asecs.org

影响力报告： 发现 Gale-ASECS 奖学金怎样帮助 学者实现充实的 研究之旅

阅读其中一位奖学金获得者如何使用数字人文工具追溯十八世纪传教书信的影响



Daniel Watkins

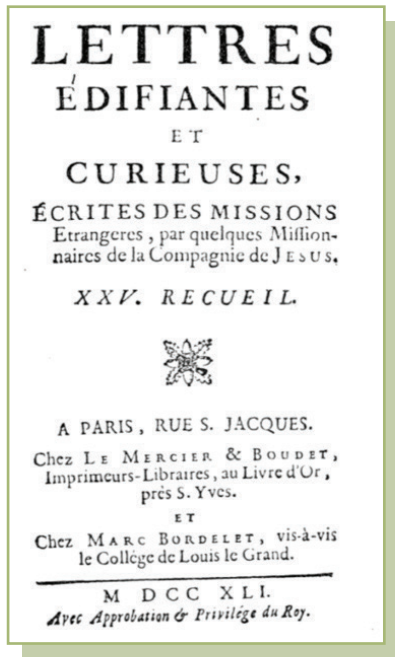
美国贝勒大学历史系助理教授，Gale-ASECS 奖学金学者

认识学者

作为 2022 年获得 Gale 和美国十八世纪研究学会（ASECS）奖学金的五位研究者之一，Daniel Watkins 使用了数字人文（DH）工具提升他对十八世纪英国《启迪和求知书信》（Lettres Édifiantes et Curieuses）的研究。在他的研究中，Watkins 希望能够发现传教书信如何影响了欧洲人对非洲、亚洲、美洲等地社会和文化的描述。

所有的 Gale-ASECS 奖学金获得者都获得了 Gale “十八世纪作品在线”（Eighteenth Century Collections Online, ECCO）和 Gale 数字学术实验室（Gale Digital Scholar Lab, 简称“实验室”）六个月的使用权，用以开展研究。ECCO 是最具综合性的在线历史档案，支持十八世纪研究，让研究者能够访问到 1701 至 1800 年之间英国印刷出版的每一部重要英文和外文图书。ECCO 中囊括的海量内容与消除了数字学术入门障碍、业界领先的文本和数据挖掘平台 Gale 数字学术实验室的巨大研究潜力结合在了一起。

学者们经过培训，掌握了实验室中提供的文本和数据挖掘工具的使用方法，以数字人文的研究方法探讨 ECCO 并推动他们的研究工作。阅读本文，了解 Watkins 是怎样利用这个基于云计算的分析平台来构建可靠的数据集，并使用数字人文工具运算得到有意义的分析结果，从而做出新发现的。



为卓有成效的研究奠定坚实的基础

Watkins 在 Gale-ASECS 奖学金期间开展的研究是一项更大研究的起点，这项研究探讨一组名为《启迪和求知书信》（Lettres Édifiantes et Curieuses, LEC）的十八世纪传教士信件的影响。“LEC 可能是最广为人知的十八世纪欧洲传教士书信典藏”，Watkins 说。“无数的同时代文献以这样或那样的形式提到这些信件，历史学家一直以来认为 LEC 有着广泛的读者群。”

Watkins 研究的第一个项目名为“追踪十八世纪英国的《启迪和求知书信》”（Tracking the Lettres Édifiantes et Curieuses in Eighteenth-Century Britain），旨在准确确定这些传教书信的影响力传播范围有多远。Watkins 指出，“这一研究的目的是帮助解决一个较大范围的历史问题，即传教士书信发挥了怎样的作用，影响了欧洲人看待世界其他地方的眼光。在这一步骤中，我着眼于十八世纪的英国，揭示 LEC 中的字句在英语文献中被再次使用和引用的证据。”

“提出一个例如‘传教士书信怎样影响了欧洲人对世界其他地方的想法’这样宽泛的问题，需要一种扩展性的研究方法。这类问题非常适合于采用数字工具开

展研究，” Watkins 承认。“实验室与 ECCO 相结合让我能够‘遥读’到 LEC 对英文出版文献的影响。特别要说的是，我关注这项奖学金正是因为 ECCO 的综合性。”

“ECCO 本身就是一套宝贵的资源，覆盖范围极广。可以说没有更好的数字资源能让我们访问到完整的十八世纪英国印刷文化。”

Taxonomy #1: Title	Taxonomy #2: Senders	Taxonomy #3: Recipients	Taxonomy #4: Place Names (Original)	Taxonomy #5: Place Names (U)
Curious	Arriot	Arendts	Agadna	Agadna
Edifying	Arieta	Armendáriz	Alep	Alep
Letters	Armendáriz	Aubers	Amerique méridionale	Bengal
	Attiret	Baltus	Anse	Buenos Aires
	Baborier	Bimet	Antoura	China
	Barbier	Brisson	Aour	Cap Haitien
	Benoist	Broissia	Arancoupan	Carataka
	Bouchet	Bruno Morales	Atipakam	Cochin China
	Bourgeois	Cibot	Balassor	Cochinchina
	Bourzes	Cochet de Saint-Vallier	Ballabaram	Conception in Chili
	Bouvet	Colittogon	Ballapouram	Ethiopia
	Burgés	Colonia	Bassora	Isfahan
	Calmette	Soudé	Bengale	Loango
	Cantova	Crécy	Buenos ayres	Guyana
	Cazier	Croiset	Cai-lum fou	Pieu
	Challier	Disabenton	Carrien-naiket-patty	Manilla
	Chansourame	Davaugour	Carton	Madurai
	Chavagnac	Delmas	Cap	Mascartines
	Choleneq	Dez	Capinagati	New France
	Chomé	Du Chambe	Careical	New Orleans
	Cibot	Du Halde	Carnate	New Philippines
	Clain	Du Trévou	Carouvepondi	Quebec
	Coeurdoux	Dugad	Cascaaskias	Tibet
	Contancin	Fleuriau d'Armenoville	Cayenne	Congo
	Crossard	Fontaney	Chandernagor	
	Dentrescolles	Germon	Chine	
	Desideri	González	Chinnaballabaram	
	Dusse	Goville	Cho tcheou	
	Dollères	Grassi	Chuchsnabouram	
	Dolu	Hibert	Cochinchine	

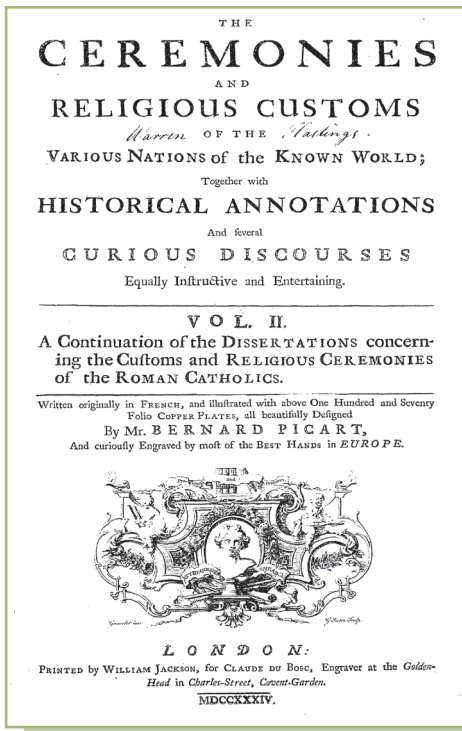
识别 LEC 使用情况的词语分类表

利用数字解决方案克服难题

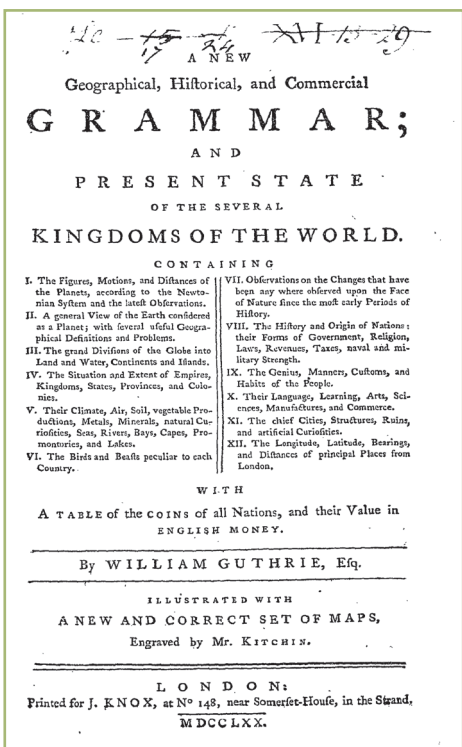
Watkins 在他研究第一阶段面临的一个直接挑战就是 LEC 的原版是法语的。Watkins 指出，“来自文献的基本关键词对于检索一个英文数据库来说是不足的。因此，我设计了一种方法，创建了一个词语分类表，其中的词语在法语和英语文献中都会出现。这个分类表包括信件作者姓名、地名以及其他从法语翻译为英语的关键词。”

“为了构建这个分类表，用以发现引用 LEC 的英文文献，我首先着眼于三十四卷的 LEC 法语版……可在‘萨宾史料’（Sabin Americana）中访问，”Watkins 说。“我使用实验室能够生成五个分类，每个分类包含 3 到 128 个词语，用以辨别 LEC 的使用情况。”





《已知世界各个国家的仪式和宗教习俗》，第2卷，Bernard Picart 著，1733年，ECCO



《新地理、历史和商贸语法》，William Guthrie 著，1770年，ECCO

获取一组可靠的数据

在建立分类表后，Watkins 概述道，“项目的第二阶段需要使用这个分类表识别引用、使用或复制了 LEC 的英文文献。使用实验室并在 ECCO 中进行检索，我设法识别出 270 部提及 LEC 的作品。其中的两部是 LEC 的英文版，其余的每部作品都是原创著作，仅出于自己的目的使用了来自 LEC 的材料。这是一项重要发现，帮助量化评价 LEC 对英国文学出版造成了多大的影响。”

谈及他在实验室中找到的文献的更多具体情况，Watkins 指出，“在 ECCO 中最重要的发现之一包括 LEC 部分卷的英译本、十八世纪欧洲最著名畅销书《已知世界各个国家的仪式和宗教习俗》的英译本，很多其他法语著作（包括伏尔泰的作品）的英译本，以及地理著作，例如《新地理、历史和商贸语法》。”

“在识别出使用 LEC 的英文作品后，我再次探讨了语料库，想要确定 LEC 是怎样被挪用以及为什么会被挪用。为了实现这个目的，我使用了实验室里的多种工具。分析结果让我初步看到 LEC 是怎样被用在十八世纪英国文学文化中的。”



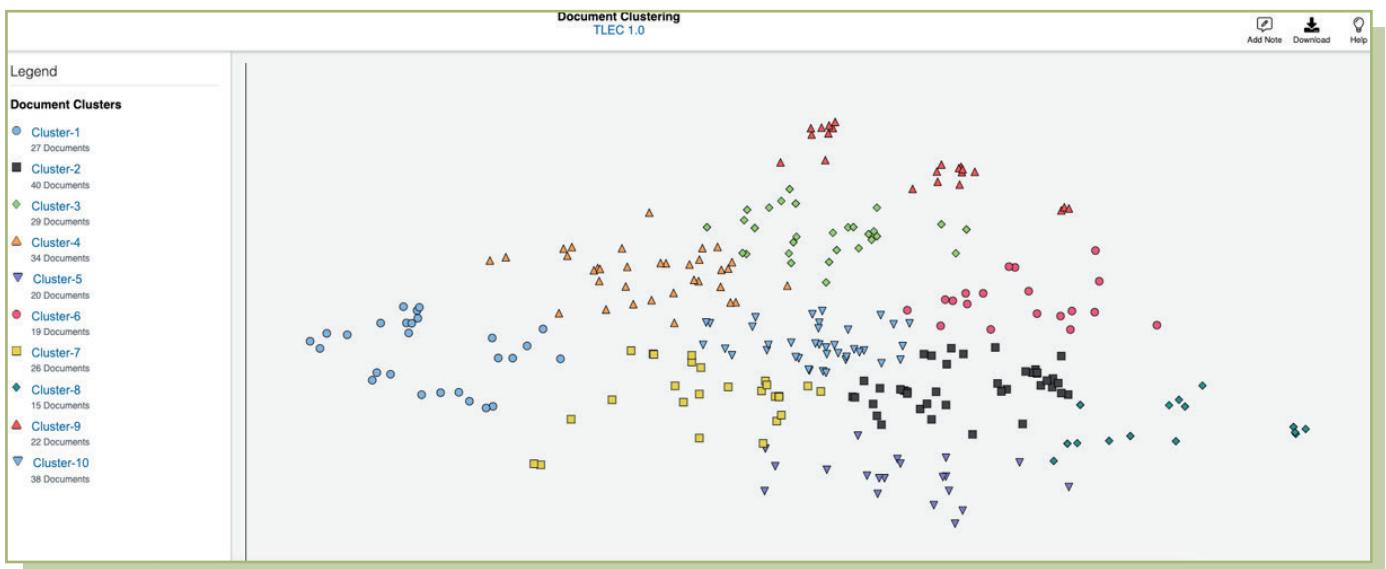
利用数字人文工具分析和可视化数据

尽管 Watkins 之前曾使用过实验室，包括在他的研究生讨论会上使用这套资源开展教学，但 Gale-ASECS 奖学金的培训让他更深入了解到如何利用数字人文研究方法拓展他的研究。“和奖学金学者们一起的培训课程非常有用，启发了我能够怎样利用实验室的想法，提供了实际操作指导，发现有效利用这些工具的方法，” Watkins 说。

当具体谈到实验室怎样提升了他的研究体验时，Watkins 说，“实验室的检索功能……是首先使用且可能是最常用的 Gale 资源。而且实验室的内容集功能让组织这些资料变得容易且有效。我构建了很多个内容

集并使用了实验室的多个功能编辑和快速研读这些资料，以了解哪些作品与 LEC 有关。此外，我使用了实验室分析工具中的四个来更好地理解语料库：文档聚类工具（clustering）、N 元语法分析工具（ngram）、主题建模工具（topic modeling）和情感分析工具（sentiment analysis）。”

“没有实验室我的研究将无法实现。就我所知没有其他工具能够像 Gale 的工具一样，让我探讨 LEC 对英语语言文学的影响这个问题。使用实验室和 ECCO 是实现我研究目标的绝对必要条件。”



Watkins 在实验室中使用文档聚类工具

利用文档聚类工具顺畅地揭示见解

在他最初的分析步骤中，Watkins 使用了文档聚类工具获取对内容集内容范围的理解。“我很好奇地想要看看是否仅是在少数文档中或基本属性相同的文档（例如同一作品的多个版本）中出现大量的 LEC 字句，” Watkins 说。“事实上，文档聚类分析揭示出非常均匀的分布。在一次识别出 10 个聚类的分析中，没有一个聚类拥有超过 40 篇文档。而且每个聚类里文档的距离也并不是特别紧密。”





Watkins 在实验室中使用主题建模工具

使用主题建模工具分析内容

Watkins 随后使用了主题建模工具可视化数据集的内容。“首先，包含在大部分文献（准确来说是 262 篇）中的主题是由单词 man（男人）、god（上帝）、people（人民）、father（父亲）、church（教会）、manner（举止）、religion（宗教）定义的。看起来这些文献似乎主要关注宗教，可能是福音传道，“Watkins 指出。“考虑到 LEC 本质上是关于耶稣会传教士福音传道努力的文献，这一点并不特别令人吃惊。但是，这证明了英国作家们并没有简单删除宗教叙事，他们在重新讲述或重新使用 LEC 时保留了这一部分。”

“第二大主题是由 trade（贸易）、aid（援助）、goods（商品）这样的单词定义的，” Watkins 说。“虽然这可能也不令人惊讶，但还是有助于证实传教士信件是对商人很有用的信息来源这一事实。重点讨论贸易的文献使用了来自 LEC 的信息，可能是告诉商人他们在长途贸易活动中会与什么样的人来往。”

最后，Watkins 指出，“第三大主题更为有趣，主要的关键词包括 king（国王）、army（军队）、王子（prince）、

general（将军）、troops（部队）、war（战争）。”Watkins 解释说，虽然对于常常出现的关键词有很多种可能的解释，但“对于潜在的传教士或宗教领袖和 / 或商人，在中国、印度、非洲和美洲这些地方的战争主题无疑是相关的。这个主题的‘独有性’分值仅为 0.318（所有主题中第四低），意味着战争是很多文献谈到的话题——其中很多可能主要关注福音传道或贸易。”

“这对我来说仅是一个开始。我还要做很多的工作来回答传教士书信如何影响了欧洲人对世界其他地方的看法这个更大的问题。数字工具将是我每一步研究不可或缺的部分。”

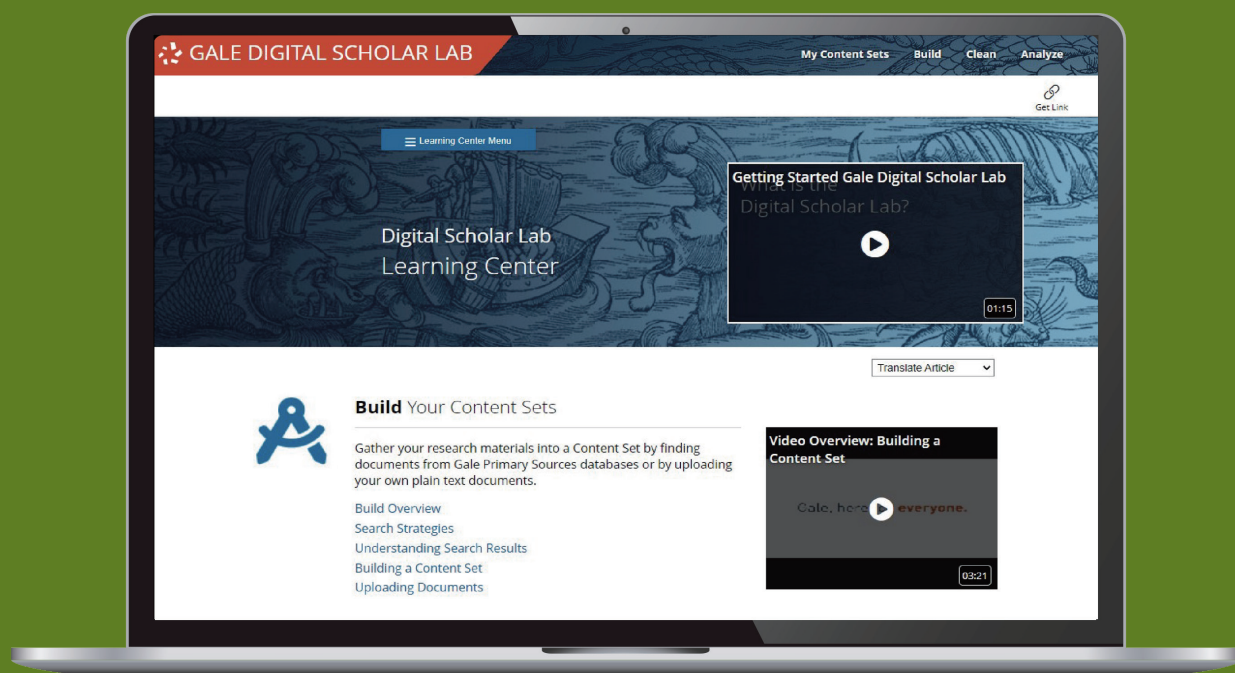
为历史学家提供对实验室潜力的深入观察

当回顾他作为 Gale-ASECS 奖学金学者的经历时，Watkins 说，“作为奖学金学者这段时间对我帮助极大，我在实验室中探索各种工具，思考这些数字工具怎样能够帮助历史学家们更好地理解过去。”

对于其他想要探索实验室中提供的数字人文工具潜力的历史学家们，Watkins 推荐道，“这些工具本身并不提供最终的结论——它们可以帮助历史学家处理信息。例如，N 元语法分析工具帮助我确定需要加入文本清理方案的额外停止词，这样主题建模和文档聚类分析就会生成更好的结果……我以前从未想过要这样使用数字人文工具。”

“这是一个绝佳的机会，我很感谢 Gale 和 ASECS 支持我的研究。当我最终发表我的研究成果时，我更是会感谢 Gale 每一个给予我帮助的人。”





查看更多有关 Gale 强大研究工具的详情

想要对 Watkins 用来提升研究水平并培养数字人文技能的 Gale 资源有更多的了解？

发现 Gale 数字学术实验室带来的各种可能性：

gale.com/digital-scholar-lab

找到更多有关 ECCO 的信息：

gale.com/ECCO

想要更多了解美国十八世纪研究学会？

AMERICAN
SOCIETY FOR
18TH-CENTURY
STUDIES

发现更多有关 ASECS 的详情，包括学会的项目计划和活动等，请访问：asecs.org

影响力报告： 发现 Gale-ASECS 奖学金怎样帮助 学者打开新的研 究路径

阅读其中一位奖学金获得者如何在 18 世纪女性作家研究中节省了大量的时间



Heather Heckman-McKenna

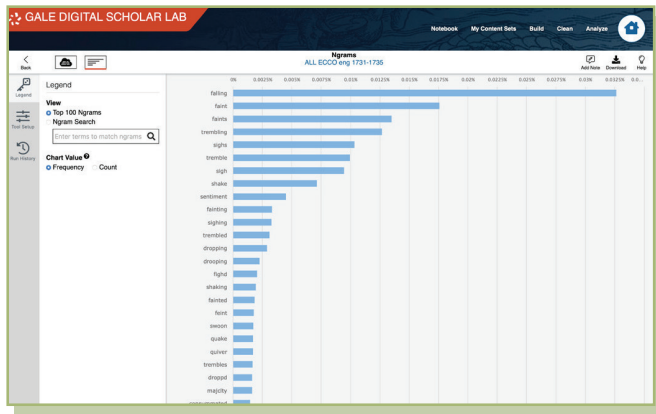
美国密苏里大学英语系博士研究生，Gale-ASECS 奖学金学者

认识学者

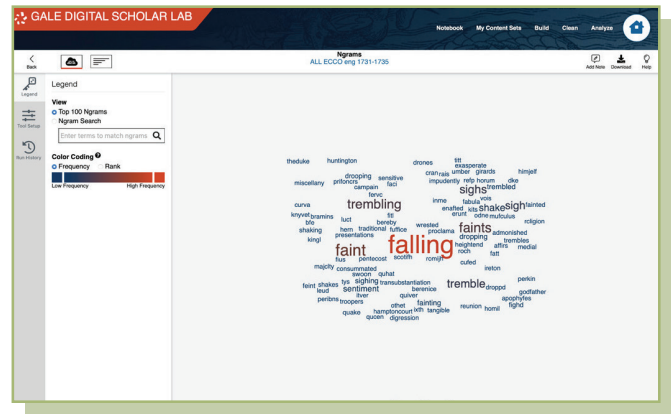
作为 2022 年获得 Gale 和美国十八世纪研究学会（ASECS）奖学金的五位研究者之一，Heather Heckman-McKenna 使用数字人文工具丰富了她对十八世纪女性作家的研究。Heckman-McKenna 开展的第一个完全基于身体破坏性动作的情感研究，旨在揭示女性作家如何利用女性动作，例如叹气、颤抖和晕厥等，营造混乱场面，作为对体制性压迫的文学回应。

所有的 Gale-ASECS 奖学金获得者都获得了 Gale “十八世纪作品在线”（Eighteenth Century Collections Online, ECCO）和 Gale 数字学术实验室（Gale Digital Scholar Lab, 简称“实验室”）六个月的使用权，用以开展研究。ECCO 是最具综合性的在线历史档案，支持十八世纪研究，让研究者能够访问到 1701 至 1800 年之间英国印刷出版的每一部重要英文和外文图书。ECCO 中囊括的海量内容与消除了数字学术入门障碍、业界领先的文本和数据挖掘平台 Gale 数字学术实验室的巨大研究潜力结合在了一起。

学者们经过培训，掌握了实验室中提供的文本和数据挖掘工具的使用方法，以数字人文的研究方法探讨 ECCO 并推动他们的研究工作。阅读本文，了解 Heckman-McKenna 是怎样利用这个基于云计算的分析平台来构建可靠的数据集，并使用数字人文工具运算得到有意义的分析结果，从而做出新发现的。



N 元语法数值



N 元语法可视化图形

将文本和数据挖掘的威力可视化

Heckman-McKenna 想要研究十八世纪文学作品中女性是怎样描写身体的。她的目标是利用 ECCO 中的文献，对那一时期文学作品中包含的叹气、颤抖、晕厥及相关的同义词文本进行编目。

在她成为 Gale-ASECS 奖学金学者之前，Heckman-McKenna 的研究受限于她所能使用的工具。她说，“我曾经检索 ECCO 找到关键词，然后手工将数据输入到 Excel 表格中。”

“实际上就在一瞬间，实验室让我立刻就能访问和下载数据，这些工作如果手工完成的话可能需要我花费几年的时间。”

Heckman-McKenna 能够充分利用实验室中的文本和数据挖掘工具，高效的管理信息并得到一幅“更加完整的画面，女性作家怎样描写感性的身体。”她说，“我知道实验室能够让我快速找到并下载大量与我的研究项目相关的数据，因此让我能够构建我的数据集。我特别感到激动的是，我可以使用实验室的 N 元语法（ngram）分析工具，确定哪些与我的研究相关的词语（例如 trembling、quivering、quaking、shaking 等）在这整段时期内或某些特定的时间段内出现的最为频繁。”

组建一套可靠的数据

那么 Heckman-McKenna 能够利用数字人文工具收集和整理哪些数据呢？“实验室让我能够找到 ECCO 中所有包含 sighs（叹气）、trembles（颤抖）和 faints（晕厥）（以及重要的同义词，例如 quivering、quavering 等）的文献，数量总计达到 51,308 条，”她说。“考虑到我寻找的是表现情感的身体动作的出现趋势，某些文献与我的研究项目并不相关。但令人高兴的是它们将与我下个阶段的工作相关：我将使用这些来自实验室的数据确定具有高度代表性的情感文字，而遥读工作将让我能够专注于精读特别具有代表性的个别文献。”

当谈到她的数据集在支持她未来研究中的潜力时，Heckman-McKenna 说，“仅仅是能够访问这些数据就是突破性的。利用下载的 .csv 和 .json 文件，我能够以几乎任何一种形式组织和分析有关身体情感的数据。例如，我可以按年代整理所有 51,308 条结果，确定关键词使用频次随时间的变化；我可以找出哪些文献包含叹气、颤抖和晕厥，而哪些仅包含颤抖和晕厥；我可以按年代仅检索颤抖的同义词，看看它们在十八世纪的不同年代出现频率怎样变化。”

“每篇文献都有如此多的数据点可供下载，这让我能够在较大的范围内进行卓有成效的调研，帮助我从宏观和微观上看到作家们是怎样描写身体情感的。我特别感谢这项奖学金给予我的这次机会。”



Text	Author	Date of Pu	Genre	sigh	tremble	Faint	HHM Code	Text
A retir'd friendship, to Ardelia	Philips	1651	Ballad			1	3	The Collected Works of Katherine Philips, p. 97
On the death of my first and dearest child	Philips	1655	Ballad			1	3	The Collected Works of Katherine Philips, p. 220
A retir'd Friendship, to Ardelia	Philips	1664	Ballad			1		British Women Poets of the Long Eighteenth Cenntury, p. 304
Orinda upon little Hector Philips	Philips	1667	Ballad			1		British Women Poets of the Long Eighteenth Cenntury, p. 678
Oroonoko		1688	Fiction	24		5	7	1 Kindle
To My Young Lover	Barker	1688	Sonnet	1				British Women Poets of the Long Eighteenth Cenntury, p. 25
Some Reflections Upon Marriage	Astell	1700	Essay			2	1	3 https://babel.hathitrust.org/cgi/pt?id=du1.ark:/13960/t3jw97d7r
A Serious Proposal to the Ladies	Astell, Mary	1701	Essay	6		1	3	T200824
The shepherds jubilee, or, A pastoral we	Smith, Dorothy	1701	Poem	1				T181046
Antiochus the great: or, the fatal relapse	Wiseman, Jane	1702	Play	7		4	3	T112280

感性行为数据库

发现新的研究路径

实验室最重要作用是为 Heckman-Mckenna 的研究提供了更为有效的路径，但也帮助提高了她的数字人文技能，发现新的研究路径。“学习主题建模（topic modeling）、词性（part of speech）分析和命名实体识别（named entity recognition）的过程中我不禁希望我能再有一年的实验室访问权，”她说。“我以前从未使用过情感分析（sentimental analysis）工具，尽管明白它更适用于现代词汇，但仍想知道它的工作原理以及我通过它能够有怎样的发现。类似的，我也很想知道聚类（clustering）工具和主题建模会帮助我在我的语料库中找到怎样的主题。”

在思考她未来的研究时，Heckman-McKenna 指出，“我现在使用的 51,038 部文献组成的数据集为我今后多年的研究提供了基础。”她特别说道，“邻近运算符检索对我而言是革命性的。比如，我可以探讨多少篇文献中叹气一词后面十五个单词之内出现了颤抖一词（或 sigh* 与 subver* 在五十个单词距离内同时出现的频率等等），这将能够为未来几年的情感研究带来重大突破。”

“我开展的研究，只有在利用实验室即刻访问到数据的情况下才可能实现。这是一套出色的工具，我很感谢能够拥有实验室的访问权限，令我的工作有了巨大的飞跃。”

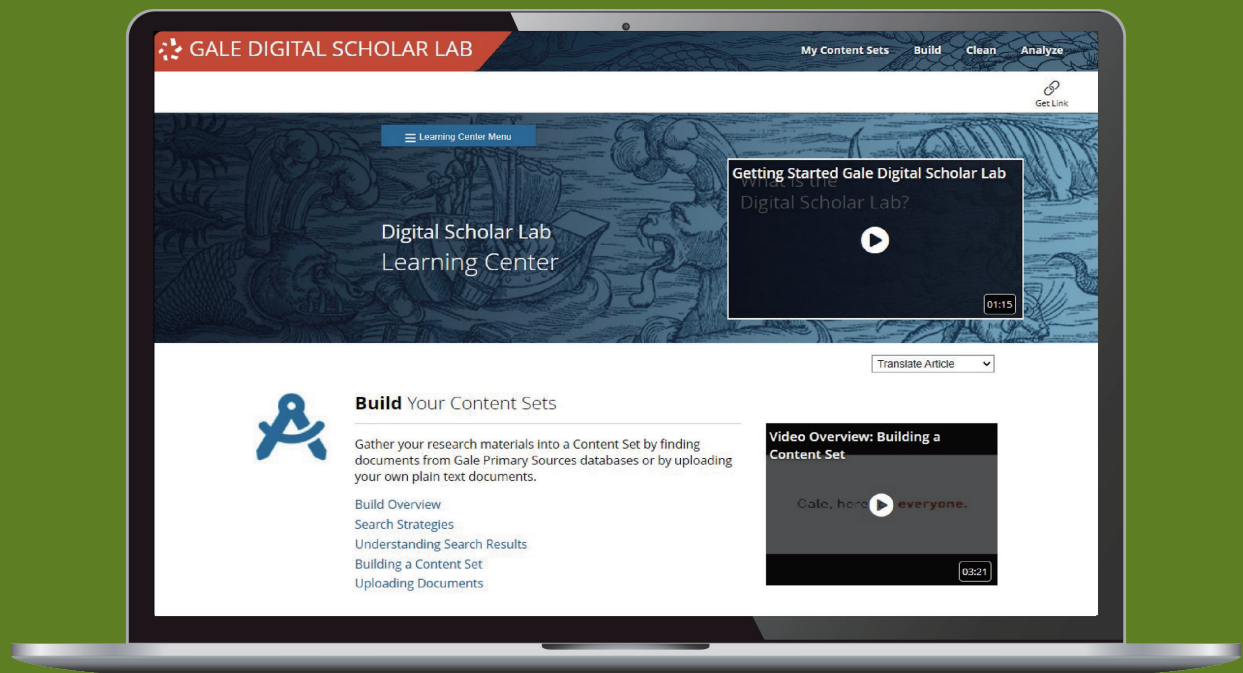
对想要了解更多的学者和教师的建议

对于其他想要采用实验室中的数字人文研究方法的学者，Heckman-McKenna 指出，“我会推荐所有实验室提供的工具。对于想要研究语言规律的学者，词性标注工具将特别有用。对于想要确定一大组数据主题的学者，我会建议使用主题建模工具。命名实体识别工具可以成为想要从地理上追踪 ECCO 数据集的学者的一个起点。简而言之，我看到了使用实验室目前提供的各项工具可能带来的巨大潜力。”

“实验室是一项不可思议的资源，显然它在今后 i 将不断改进。Gale 已经开发的下载和分析功能为学者提供了重要的机会，找到新的方式对十八世纪文学提出问题并进行思考。”

对于想要在教学中分享数字人文研究方法的教师，Heckman-McKenna 也推荐道，“在本科生教学中，我想我会更多地使用邻近运算符检索和 N 元语法。让学生能够根据 N 元语法结果自由地提出自己的研究问题，这将让学生们意识到文学学习远不只是精读书本。他们将了解到他们可以用更有条理、更科学的思维研究文学，发现潜在的全新方式思考文献。”





查看更多有关 Gale 强大研究工具的详情

想要对 Heckman-McKenna 用来提升研究水平并培养数字人文技能的 Gale 资源有更多的了解？

发现 Gale 数字学术实验室带来的各种可能性：

gale.com/digital-scholar-lab

找到更多有关 ECCO 的信息：

gale.com/ECCO

想要更多了解美国十八世纪研究学会？

AMERICAN
SOCIETY FOR
18TH-CENTURY
STUDIES

发现更多有关 ASECS 的详情，包括学会的项目计划和活动等，请访问：asecs.org

影响力报告：

发现 Gale-ASECS 奖学金怎样帮助 学者启迪新的研 究问题

阅读其中一位奖学金获得者如何使用数字人文工具理解十八世纪语言和交流障碍的表达



Jared Richman

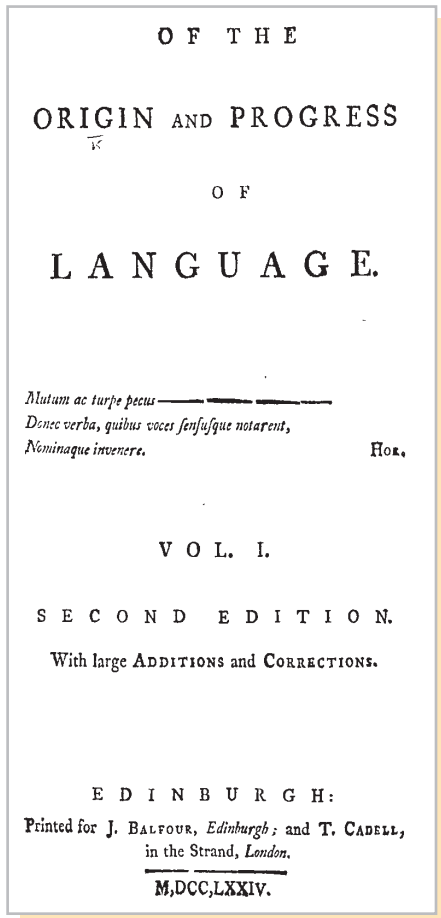
美国科罗拉多学院英语系副教授，Gale-ASECS 奖学金学者

认识学者

作为 2022 年获得 Gale 和美国十八世纪研究学会 (ASECS) 奖学金的五位研究者之一，Jared Richman 使用数字人文工具提升他对十八世纪作家如何表达和政治化语言和交流障碍的研究。Richman 的研究项目“十八世纪印刷文化中的发音障碍” (Voicing Disability in Eighteenth-Century Print Culture) 基于该领域内其他残疾学者的研究工作而设计。

所有的 Gale-ASECS 奖学金获得者都获得了 Gale “十八世纪作品在线” (Eighteenth Century Collections Online, ECCO) 和 Gale 数字学术实验室 (Gale Digital Scholar Lab, 简称“实验室”) 六个月的使用权，用以开展研究。ECCO 是最具综合性的在线历史档案，支持十八世纪研究，让研究者能够访问到 1701 至 1800 年之间英国印刷出版的每一部重要英文和外文图书。ECCO 中囊括的海量内容与消除了数字学术入门障碍、业界领先的文本和数据挖掘平台 Gale 数字学术实验室的巨大研究潜力结合在了一起。

学者们经过培训，掌握了实验室中提供的文本和数据挖掘工具的使用方法，以数字人文的研究方法探讨 ECCO 并推动他们的研究工作。阅读本文，了解 Richman 是怎样利用这个基于云计算的分析平台来构建可靠的数据集，并使用数字人文工具运算得到有意义的分析结果，从而做出新发现的。



《语言的起源和发展》，James Burnet, 1774 年，ECCO。

以残疾学者的批判性见解为基础

在作为 Gale-ASECS 奖学金学者期间，Richman 的目标是深入探究十八世纪语言和交流障碍的各种相关评述。在总结他的研究时 Richman 说，“这个研究项目描绘出十八世纪，以及这个世纪关于人性准确定义的问题，怎样体现出和政治化言语及非规范化的语言表达。”

他进一步解释说，“我理解语言及其发音是一个争夺的空间，对于不懈定义人性的努力至关重要。这种争夺从近代早期一直延续至今。因此，通过汇集朗诵的历史及近代早期背景下的残疾人论述，我拓展了残疾学者们的研究，这些学者例如 Lennard Davis、David Turner、Christopher Mounsey、Felicity Nussbaum 和 Helen Deutsch，以及语言和语音历史学家，例如 Paula McDowell、Jack Lynch 和 Janet Sorensen。”

建立研究框架

甚至在开始查找数据之前，Richman 就需要进一步定义他要寻找的内容的类型。他指出，“要讨论语言（以及与语言相关的残疾），我们必须从划定我们的问题范围开始。这并非一项简单的工作，因为涉及到漫长十八世纪以及常常与这个世纪相关联的启蒙运动时期的态度和行动。也许因为印刷文化的爆发式发展，这一时期特别关注发生语音交流的各种不同方式。Grumble（嘟囔）、mumble（咕哝）、ramble（絮叨）、maunt、mant、babble（唠叨）、burble（嘟囔）、blurt（脱口而出）、blather（喋喋不休）、chatter（唠叨）、flatter（奉承）、lisp（口齿不清）、sputter（噼啪声）、stutter（结结巴巴）、stammer（结巴）、prattle（唧唧呀呀）、tattle（闲聊）、utter（说出）、whisper（耳语）、ejaculate（突然说出）、groan（叹息）、moan（呻吟）：为了宣布一件事情是怎样的，十八世纪的英国人继承、发明或启用了大量的词语来描述语言的模式或形式。

“因此，我研究项目的第一个重要问题是，在十八世纪印刷文化中，我们怎样以及在那里能够找到对声音残疾的描写？”Richman 概括说。“一些学者曾认为（尽管证据表明是相反的），‘残疾’在十八世纪还不算是一种个人身份和体现（因为这个词在十九世纪中期才具备了现代的定义）。但我认为，这个词不能定义一个固定的类别并不意味着近代早期文献中没有对残疾的表述——远非如此。”

Richman 解释说，“近代早期强调语言的正确性和标准化，实际上‘残疾’的概念在这个时期恰恰正如我们所知，在众多学科（医学、哲学、戏剧、语言学、美学、诗歌等）的话语中被固定下来。我在这里想要做的是，使用语音残疾和交流障碍作为框架，找到 disability（残疾）一词以外的词语，发现他们存在于哪里，出现在哪些语境和体裁中，它们在这些具体的实例中发挥怎样的作用，以及它们如何指向我们如今称之为‘残疾’的现代概念。”



影响力报告：

发现 Gale-ASECS 奖学金怎样帮助学者启迪新的研究问题

Document Title	Con ent Type	Document Types	Publication Title	Publication Date	Publisher	Place of Publication
Hymns. By the Rev. S. Medley, of Liverpool	Monograph	Monograph	Religion and Philosophy I.	1789	Gale	London, United Kingdom
The history of Julia D'Haumont: or the eventful connection of the house of Montellian with that of D'Haumont. By the Rev. James Douglas, of Chiddingfold, Sussex. Vol. I	Monograph	Monograph	Literature and Language I.	1797	Gale	London, United Kingdom
Prolegomena in versione syriacâ philoxenianam Novi Testamenti. Editi J. White, S.T.P. Ling. Arab. Apud Oxonienses Prof.	Monograph	Monograph	Religion and Philosophy II.	1798	Gale, part of Cengage Group	Oxford, United Kingdom
Detached pieces, in verse; by Thos. Canning, Private Soldier, 61st. Regt.	Monograph	Monograph	Literature and Language I.	[1792?]	Gale	London, United Kingdom
A serious alarm to the people of Bath; or, the diabolical soliloquy of an old witch. Overheard in the Crescent, by an apothecary of Bath	Monograph	Monograph	Literature and Language I.	MDCCLXXXII. [1782]	Gale	London, United Kingdom
The pittiad: A satire. By Do! Common	Monograph	Monograph	Literature and Language I.	MDCCLXIX. [1769]	Gale	London, United Kingdom
The age of infidelity: part II. In answer to the second part of The age of reason. With some additional remarks upon the former. By a layman, the second edition	Monograph	Monograph	Religion and Philosophy I.	1796	Gale	London, United Kingdom
An historical account of the life and reign of David King of Israel. In four books. ... In which ... Mr. Bayle's criticisms, ... are fully considered. By the author of Revelation examined with candour. In two volumes.	Monograph	Monograph	Religion and Philosophy I.	1759	Gale	London, United Kingdom
A new chronological abridgement of the history of France, Containing The Publick Transactions of that kingdom, from Clovis to Lewis XIV. their wars, battles, sieges, &c. their laws, manners, customs, &c. Written	Monograph	Monograph	History and Geography I.	MDCCLXII. [1762]	Gale	London, United Kingdom
By permission of the Right Worshipful the Vice-Chancellor, and the Worshipful the Mayor. At the Town-Hall, on Wednesday evening, the 20th of June instant, at half-past six o'clock. Captain S. Villiers has the hon	Monograph	Monograph	Social Sciences I.	[1789?]	Gale	London, United Kingdom
An evening conversation	Monograph	Monograph	Literature and Language I.	MDCCLXXXVII. [1737]	Gale	San Marino, CA, United States
Ministers of the gospel witnesses for Christ. A sermon, preached before The Reverend John Carver, B.L.L. Archdeacon of Surrey, at his visitation, held in the Parish Church of St. Saviour, Southwark, October 5th	Monograph	Monograph	Religion and Philosophy I.	1796	Gale	London, United Kingdom
Rameltes: a poem, humbly inscrib'd to his Grace, the D. of Marlborough. By W. Wagstaff, Gent. of Lincoln College Oxon	Monograph	Monograph	Literature and Language I.	1706	Gale	Cambridge, MA, United States
Mad Tom; or, the bard of Bath	Monograph	Monograph	Literature and Language I.	MDCCLXXX. [1781]	Gale	London, United Kingdom
The trial of James N. Tandy, Esq, as had in His Majesty's Court of King's Bench, on the 11th of June, 1792, on an indictment for challenging and provoking to fight, John Toler, Esq, His Majesty's solicitor general. Taken in short hand, by a barrister	Monograph	Monograph	Literature and Language I.	1792	Gale, part of Cengage Group	Cambridge, United Kingdom
A grammatical history of the English language; comprising, an easy, concise and systematic method of education; designed for the use of schools in America. In three parts. Part III. Containing the necessary	Monograph	Monograph	Literature and Language I.	[1785]	Gale	London, United Kingdom
A true account of the siege of London-Derry. By the Rev. Mr. George Walker, rector of Donaghmore, in the county of Tyrone, and late governor of Derry, in Ireland. To which is added, Sir John Dalrymple's account	Monograph	Monograph	Social Sciences II.	MDCCLXXXVII. [1787]	Gale, part of Cengage Group	Dublin, Ireland
An address to the public, concerning impediments in speech, defects of utterance, &c. and a syllabus of a course of lectures on public speaking	Monograph	Monograph	Social Sciences I.	MDCCLXXXVII. [1786]	Gale	Oxford, United Kingdom
Eclectic excursions on, literary & pictorial sketches of countenance, character & country, in different parts of England & South Wales. ... Embellished with upwards of one hundred characteristic & illustrative	Monograph	Monograph	Literature and Language I.	1796	Gale	Cambridge, MA, United States
The works of Mrs. Behn. Entire in three volumes. ...	Monograph	Monograph	Literature and Language II.	1711	Gale, part of Cengage Group	Austin, TX, United States
Cerberus; or, A leash of portraits. A poem	Monograph	Monograph	Literature and Language II.	MDCCLXXXVIII. [1788]	Gale, part of Cengage Group	Cambridge, MA, United States
An American selection of lessons in reading and speaking. Calculated to improve the minds and refine the taste of youth. And also to instruct them in the geography, history, and politics of the United States. To	Monograph	Monograph	Literature and Language II.	MDCCLXXXIX. [1789]	Gale, part of Cengage Group	Washington, DC, United States
An American selection of lessons in reading and speaking. Calculated to improve the minds and refine the taste of youth. And also to instruct them in the geography, history, and politics of the United States. To	Monograph	Monograph	Literature and Language II.	...1797	Gale, part of Cengage Group	Washington, DC, United States
Poems, by S. T. Coleridge, second edition. To which are now added poems by Charles Lamb, and Charles Lloyd	Monograph	Monograph	Literature and Language I.	1797	Gale	Cambridge, MA, United States
The gardeners kalendar, directing what works are necessary to be done every month, in the kitchen, fruit, and pleasure gardens, and in the conservatory. With an account of the particular reasons for the propos	Monograph	Monograph	Science, Technology, and Medicine II.	MDCCLXXXII. [1732]	Gale, part of Cengage Group	Cambridge, United Kingdom
Moral epistle, respectfully dedicated to Earl Stanhope	Monograph	Monograph	Literature and Language I.	MDCCLXCV. [1765]	Gale	London, United Kingdom
The beauties of the English stage: consisting of the most affecting and sentimental passages, ... in the English plays, ancient and modern. ... The third edition. ... In three volumes. ...	Monograph	Monograph	Literature and Language I.	1756	Gale	London, United Kingdom
The mysterious mother: A tragedy. By Mr. Horace Walpole	Monograph	Monograph	Literature and Language II.	MDCCLXVII. [1768]	Gale, part of Cengage Group	London, United Kingdom
The ruins of Linnithgow, with selected notes in vindication of the character of Mary, Queen of Scots. By Joseph D. M'arthur, North York	Monograph	Monograph	Social Sciences II.	1798	Gale, part of Cengage Group	Oxford, United Kingdom
A caution to gentlemen who use Sheridan's dictionary. To which are added, for the assistance of foreigners and natives, select rules for pronouncing English with precision and elegance	Monograph	Monograph	Literature and Language II.	MDCCLXCI. [1781]	Gale, part of Cengage Group	Dublin, Ireland
A catalogue of rare, splendid, and valuable books, in every branch of polite literature; including the entire libraries of the Rev. Harvey Spragg, ... also of the Rev. Henry Putnam, ... The sale will begin on February	Monograph	Monograph	Literature and Language I.	1798	Gale	London, United Kingdom
The friendly and honest advice of an old tory to the Vice-Chancellor of Cambridge	Monograph	Monograph	Social Sciences I.	MDCCLJ. [1751]	Gale	London, United Kingdom
A sermon preach'd May the 4th. 1707, at Uxbridge, upon occasion of the late day of thanksgiving for the union of the two kingdoms of England and Scotland. By John Olyffe, M.A. Rector of Hedgerly Bucks	Monograph	Monograph	Religion and Philosophy II.	1707	Gale, part of Cengage Group	Oxford, United Kingdom
The brother's gift; or, the naughty girl reformed. Published for The advantage of the rising generation. Ye British Ladies, would you wish to see Abeauteous Offspring grating round your Knee, Shun each	Monograph	Monograph	Literature and Language I.	[1775]	Gale	Manchester, United Kingdom
An alarm to an inland navigation, to Ireland. Estimated in a dialogue between a gentleman in the country, and a citizen of Dublin, in spring, 1781. Occasioned by the great scarcity of coals then in the city	Monograph	Monograph	Social Sciences II.	MDCCLXI. [1761]	Gale, part of Cengage Group	Oxford, United Kingdom
This day the poll was continued for electing two Members to represent the City of Dublin in Parliament, and at the Adjournment of the Court the Poll stood thus,	Monograph	Monograph	Law I.	[1761]	Gale	London, United Kingdom
An alarm to the people of England; shewing their rights, liberties, and properties, to be in the utmost danger from the present destructive, and unconstitutional association, for the preservation of the game all o	Monograph	Monograph	Social Sciences I.	MDCCLVII. [1757]	Gale	London, United Kingdom
Demosthenous logoi Demegorikoi Dodeka. = Demosthenis orationes de republica duodecim, cum Wolfiana interpretatione denuo castigata, et notis historicis J. V. Lucchesini. Accessit Philippi epistola. Editi G	Monograph	Monograph	Religion and Philosophy I.	MDCCLV. [1755]	Gale	London, United Kingdom
The new week's preparation for a worthy receiving of the Lord's Supper, as appointed and practised by the Church of England; part the second: Consisting of Meditations, Prayers, and Hymns suitable for the S	Monograph	Monograph	Religion and Philosophy I.	1759	Gale	Lawrence, KS, United States
Unto the Right Honourable the Lords of Council and Session, the petition of Sir James Lockhart of Carstairs, baronet, ...	Monograph	Monograph	Law II.	[1747]	Gale, part of Cengage Group	Edinburgh, SF, United Kingdom
A catalogue of the valuable farming stock, implements in Husbandry, Dairy and Brewing Utensils, of the late Rev. Tobias Rustat, Deceased, At Stutton, Suffolk, which will be sold by auction, by Messrs. Harris &	Monograph	Monograph	General Reference I.	[1783]	Gale	London, United Kingdom
The mysterious mother: a tragedy	Monograph	Monograph	Literature and Language I.	1791	Gale	Manchester, United Kingdom
Letters of the late Rev. Mr. Laurence Sterne, To his most intimate Friends. With a fragment in the Manner of Rabelais. To which are prefix'd, Memoirs of his Life and Family Written by himself And published by hi	Monograph	Monograph	Literature and Language I.	1778	Gale	Oxford, United Kingdom
A poem on the happy return of His most Sacred Majesty King George. Humbly inscrib'd to His Royal Highness the Prince	Monograph	Monograph	Literature and Language I.	1719	Gale	Cambridge, MA, United States
Atys and Adrastus, a tale in the manner of Dryden's fables. By Mr. William Whitehead, Fellow of Clare-Hall in Cambridge	Monograph	Monograph	Literature and Language I.	1719	Gale	Cambridge, MA, United States
The noble game of chess; or, A New and easy Method to learn to play well in a short Time: Together with a curious Account of its Antiquity, Derivation of its Terms, &c. By Phillip Stamma, Native of Aleppo in Syria	Monograph	Monograph	Fine Arts I.	1745	Gale	London, United Kingdom
The psalm singer's guide: being a choice collection of the most useful tunes of the psalms, in two, three, and four parts, the Upper Parts being all in the G-Sol-re-ut Cliff: With Psalms and Hymns on several Occa	Monograph	Monograph	Fine Arts I.	1719	Gale	London, United Kingdom
Liberty: A poem. To the Queen	Monograph	Monograph	Literature and Language I.	MDCCLXXX. [1729]	Gale	London, United Kingdom

Richman 为 “stutter”（口吃）一词建立的文献内容集

驾驭文本和数据挖掘的力量

Richman 在成为 Gale-ASECS 奖学金学者前几年就开始了他对这个主题的研究。在奖学金期间，他拥有的相关资料和数字人文工具的访问权帮助他走上了成果更为丰盛的研究道路。

“我意识到我正在进行的研究项目已经到达了一个点，在这里我会需要定量的数据，” Richman 说。“我原来的文献查找方式（旅行前往档案馆）在 2020 年 3 月因全球疫情的爆发而中断。我在之前访问档案馆的旅行以及检索一些图书馆的数字档案中获得了大量的数据，而且我意识到我需要扩展我的一次文献内容集。虽然已经有关于十八世纪修辞法和朗诵的重要著作，但还没有方法交叉检索这些研究和我自己关于残疾的研究工作。Gale 奖学金打开了一条可能的道路，将这些研究线索汇聚在一起。”

Richman 解释说，“能够以这种方式访问 ECCO 中庞大的十八世纪印刷文献典藏，可以按年代、体裁、主题和其他特征检索和标注不同类型的文献，这真是太棒了。尽管我过去曾使用过 ECCO 来获取某些特定时间段内的文献（主要是历史、政治、医学和文学作品），我还从未根据特定的检索词分组和区分文献（当然也从没这样迅速过！）。而且，能够快速构建（和形成）文献内容集让我可以更轻松地处理大量的信息，所花费的时间只是通常所需时间的一小部分。”

“虽然各家图书馆的目录通常提供了丰富的细节，例如标题、作者、主题等，但它们极少能让研究者以这样的速度和深度挖掘模糊的检索词和名称。能够在实验室中检索和分析如此巨大数量的内容并得到有意义的分析结果，这对我的工作大有裨益。”



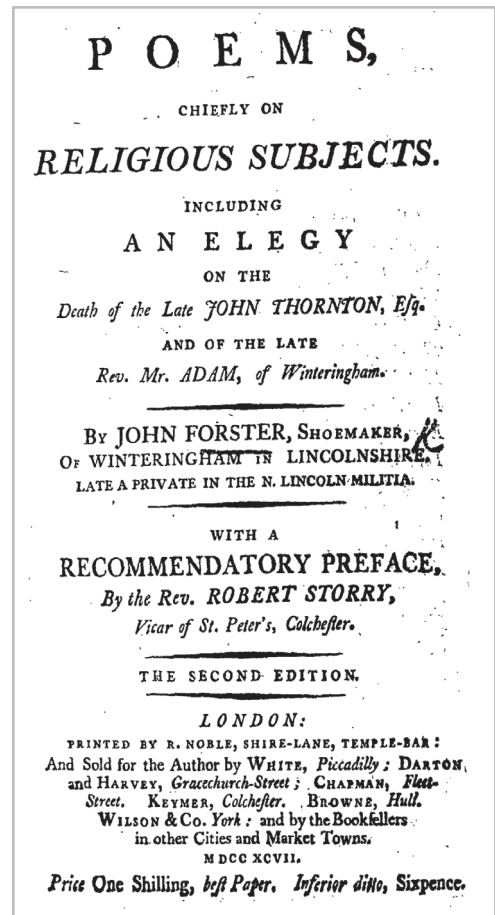
发现新的研究道路

当被问到实验室怎样以全新的方式影响了他的研究时，Richman 说，“我最惊讶的是它提出了新的研究问题。

“我在数字人文研究领域相对来说是名新手（除了在文学方面之外），我逐渐看到在研究中采用的一些算法可能产生的不仅仅是一套一套的数据，也有可能提出新的问题……能够快速分析的大量内容也为现有的研究项目打开了新的窗口，” Richman 指出。

他解释说，“例如，当我尝试找出与残疾和语音相关的趋势和聚类随时间的变化时，我看到对于与我使用的一般检索词相关的文献的体裁和形式，有很多有趣的问题值得探讨。我逐渐意识到我的研究项目不仅需要着眼于这些词语和概念各自的情况，还要考虑并行和相关的公共话语（政治思想、文学、医学）在使用残疾和非正规语言相关概念时的分歧或重叠。我仍在建立新研究问题的过程中，围绕文学形式和印刷体裁，这将构成这个研究项目的下一个阶段。”

“我由衷地感谢访问实验室的机会。如我所述，我在数字人文研究和方法学方面算是一名新手，但我非常高兴发现了新的方法建立研究问题，而这些问题已经成为了我学术研究的核心。”



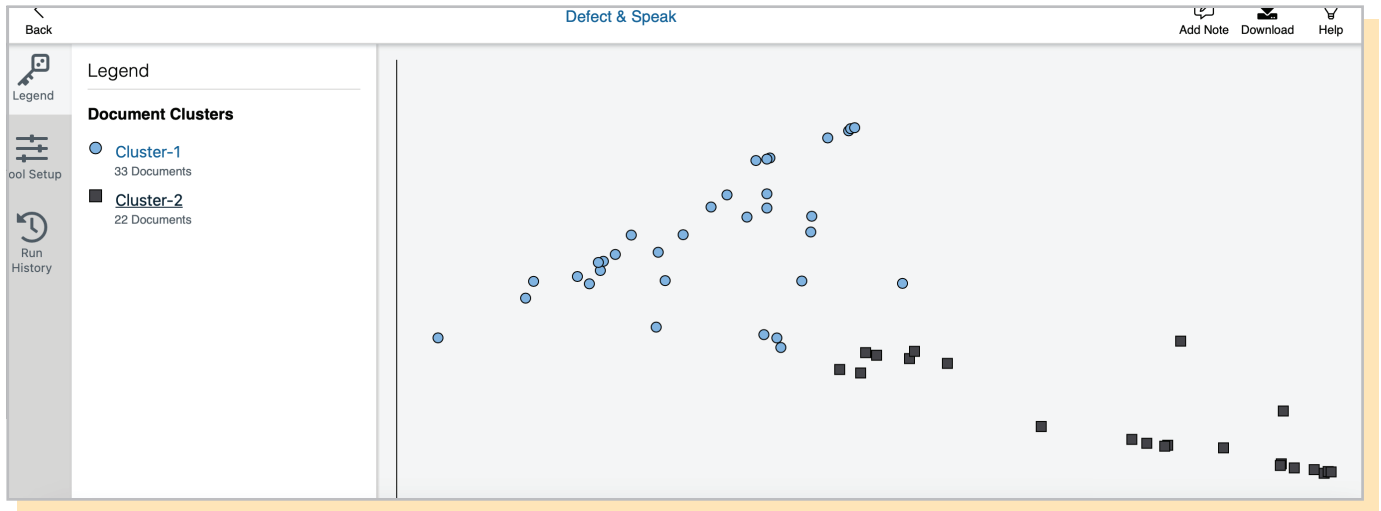
《诗歌，主要关于宗教主题：包括致已故约翰·桑顿的挽歌》，John Forster, 1797 年，ECCO。

发现出乎意料的相关叙述

那么实验室令 Richman 发现了哪些类型的内容呢？

“我利用实验室在 ECCO 中发现的一套有趣的文献在军事背景下对残疾概念进行了分组，” Richman 说。“一个例子是 John Forster 著《诗歌，主要关于宗教主题：包括致已故约翰·桑顿的挽歌》（Poems, Chiefly on Religious Subjects. Including An Elegy on the Death of the Late John Thornton, 1797 年）。这部作品标志着实验室出人意料的检索能力——如果我只是在图书馆目录中看到这篇文献，我很可能不会停下来阅读全文。然而，因为我能够交叉参照特定的词语，我发现一首老兵写的诗强调了他的残疾（他的身份是一名残疾退伍军人）。作品本身的文字突出了他的写作生涯（作为一名诗人）、他的军队经历和他的身体缺陷之间的关联。”





有关“defect”（缺陷）和“speak”（讲话）的文献聚类

利用数字人文工具发现新见解

总结他在实验室中分析数据集的经验，Richman 指出，“对我来说最有趣的是追踪与我的研究项目有关的特定词语在这个世纪（1700-1800 年）的出现和使用。我猜测与语音障碍相关的词语（例如 stutter、stammer、lisp、hesitate [犹豫]、mumble、mutter [嘀咕] 等）会随着英国朗诵运动的出现而增多……我没有可以量化的数据支持我的猜想。实验室让我能够收集与这些检索词以及 voice（声音）、speech（语音）、tongue（语言）和 elocution（朗诵）等词相关的数据，生成这些词语的历史地图。”

Richman 解释说，“例如，出现 elocution 一词的作品在 1750 年以后增加了两倍多（出现 voice 一词的作品也是如此）。而且，随着专门论述或出现 voice 和 elocution 的作品越来越多，defect（缺陷）、stammer、stutter 和 lisp 这些词在作品中也越来越多的出现，二者随时间出现的频率几乎一致。

“我仍在想办法解释印刷作品中的这种现象，但看起来语音正确性方面文化意识的提高恰好与对不正确或‘缺陷’语音，即十八世纪下半叶所说的残疾的关注同一时期发生，”他总结道。

为未来的研究和教学做准备

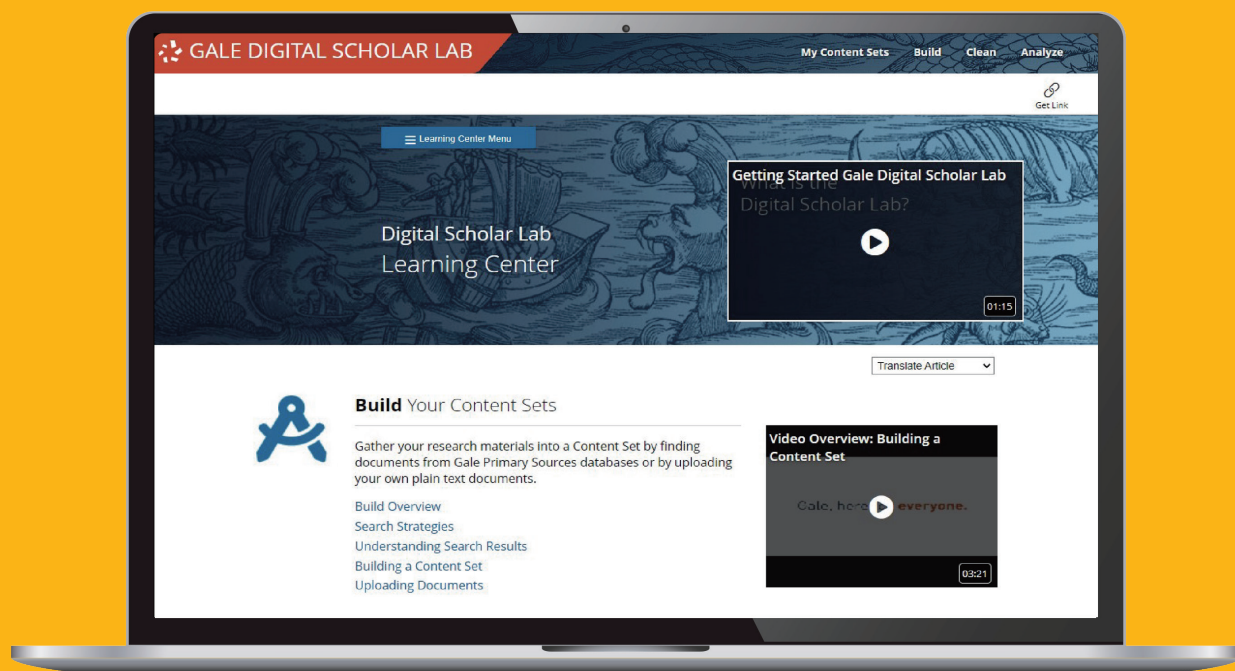
当谈及他在 Gale-ASECS 奖学金期间收集的数据的潜在应用时，Richman 说，“我将会使用我收集的数据工作一段时间，继续分析和优化这些数据并思考如何能够最好地准确呈现结果，让我想要接触到的各个读者群都能够轻松地理解。

“我计划在这一研究项目的其他方面和未来的研究项目中试验更多的数字人文方法，” Richman 说。

作为一名英语系的副教授，Richman 也指出，“我很高兴将实验室带到课堂上，和我的学生们一同使用。我很多年来一直在使用 ECCO，让我的学生阅读没有现代版本的英国早期作品，传授基本的档案学方法，研究近代早期的诗歌等。我已经看到了可以使用实验室的教学练习，其中的工具将帮助学生分析和处理 ECCO 中的文献。”

“能够有时间、空间和工具真正深入挖掘这些全新的数字方法对我的工作极有帮助。这也为我带来了新鲜感，我的研究和教学中将有各种各样数字人文的可能性。”





查看更多有关 Gale 强大研究工具的详情

想要对 Richman 用来提升研究水平并培养数字人文技能的 Gale 资源有更多的了解？

发现 Gale 数字学术实验室带来的各种可能性：

gale.com/digital-scholar-lab

找到更多有关 ECCO 的信息：

gale.com/ECCO

想要更多了解美国十八世纪研究学会？

AMERICAN
SOCIETY FOR
18TH-CENTURY
STUDIES

发现更多有关 ASECS 的详情，包括学会的项目计划和活动等，请访问：asecs.org

影响力报告：

发现 Gale-ASECS 奖学金怎样帮助 学者拓展研究可 能性

阅读其中一位奖学金获得者如何使用数字人文工具分析十八世纪数量词的使用



Sarah Weston

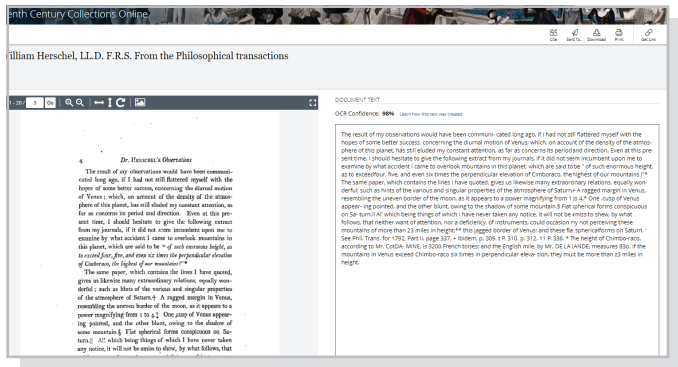
美国耶鲁大学英语系博士生，Gale-ASECS 奖学金学者

认识学者

作为 2022 年获得 Gale 和美国十八世纪研究学会（ASECS）奖学金的五位研究者之一，Sarah Weston 使用数字人文工具从数学的视角考察十八世纪的诗歌和艺术作品。Weston 的项目“数字的形状，1701-1800 年”（The Shape of Numbers, 1701-1800）分析数量词，以一种全新的方式探讨从启蒙运动到浪漫主义的社会发展，以及十八世纪末发生的情感变化。

所有的 Gale-ASECS 奖学金获得者都获得了 Gale “十八世纪作品在线”（Eighteenth Century Collections Online, ECCO）和 Gale 数字学术实验室（Gale Digital Scholar Lab, 简称“实验室”）六个月的使用权，用以开展研究。ECCO 是最具综合性的在线历史档案，支持十八世纪研究，让研究者能够访问到 1701 至 1800 年之间英国印刷出版的每一部重要英文和外文图书。ECCO 中囊括的海量内容与消除了数字学术入门障碍、业界领先的文本和数据挖掘平台 Gale 数字学术实验室的巨大研究潜力结合在了一起。

学者们经过培训，掌握了实验室中提供的文本和数据挖掘工具的使用方法，以数字人文的研究方法探讨 ECCO 并推动他们的研究工作。阅读本文，了解 Weston 是怎样利用这个基于云计算的分析平台来构建可靠的数据集，并使用数字人文工具运算得到有意义的分析结果，从而做出新发现的。



并行查看文献复本图像和可下载的 OCR 文本

利用数字人文工具提升研究

Weston 在 Gale-ASECS 奖学金期间的研究集中于她的一个大型项目，考察十八世纪的数量概念。“我目前这本著作的研究工作是探讨 1660 年至 1840 年人们对数字的态度转变，我们看到在十八世纪早期和晚期之间人们在很多不同领域内对数量的想法出现了差别，从诗歌和艺术到科学和数学，” Weston 指出。

“‘数字的形状’是这个项目的一个数字人文分支，使用（实验室）提供的自然语言处理（NLP）工具和开放获取工具 Voyant，追踪数字和数量的使用怎样在一个世纪的时间内发生变化。实验室的特别之处在于能够让我找到并研究更大型的文献样本集，比我之前使用的数据更为庞大，也能够帮助我发现我之前并不知道的文献，这些都正在成为我为我的这本著作所作的研究的一部分。”

Weston 描述了在她的研究中应用数字人文工具是一个自然而然的过程，对十八世纪的内容进行基于数字的分析。“转向数字人文显然是一个必要的步骤（从数字上评价和思考文学），考察定量方法怎样能够帮助我回答我的研究问题，”她说。

“实验室对我而言是我研究工作的一项宝贵工具，我在这个数字人文项目中所做的工作将帮助我完成我的著作，从论文转化为一本专著。”

驾驭文本和数据挖掘的力量

Gale-ASECS 奖学金为 Weston 提供了数字人文工具和培训，支持她走入研究工作的下一个阶段。

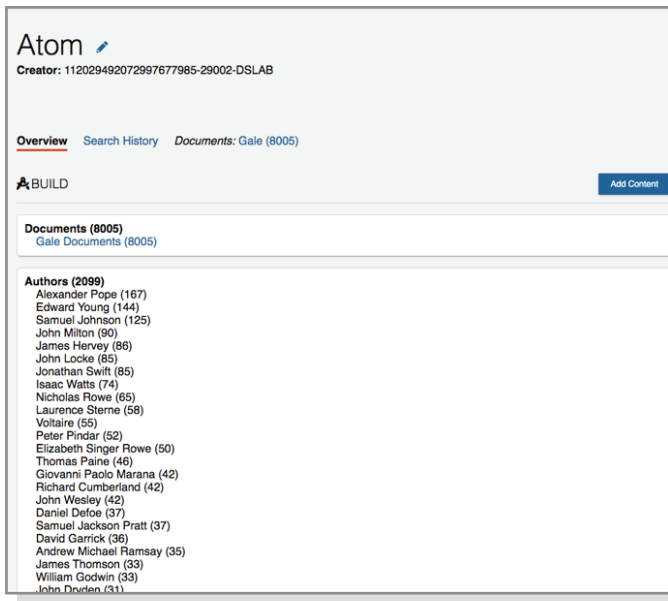
在回顾拥有实验室访问权之前的研究经历时，Weston 说，“我的论据主要围绕对关键词的精读，但因为我研究的是整个十八世纪在大范围内对数量的重新定义，我也曾尝试全方位的理解这个时间段内使用的数量词。实验室真正帮助我拓展了我对这个研究项目中包含的各种类型文献的理解，不仅仅是诗歌和哲学作品，也包括艺术专论和数学教材，以及园林设计图书、儿童读物和年鉴等……我能够发现我现有研究中的空白，找到相关的数量词应用场景填补缺口，很多都是我没有想到的或我已经忘记的领域。”

Weston 解释说，“除了帮助我创建一个更好、更大的数据集，实验室也帮助我组织、筛选和理解我正在使用的文献资料……能够直接并行查看 OCR（光学字符识别）文本和文献复本图像，对于我筛选这些新资料非常有帮助。我可以复制粘贴 OCR 文本，在我对有用的项目做笔记以便今后使用时节省了很长的时间。因此实验室的各种功能和工具及其与 ECCO 和其他我正在使用的数据库的无缝整合，对我整理和准备用于分析的文本帮助很大。”

“完成每件事情的速度之快令人惊讶，” Weston 补充说，总结了实验室是怎样帮助简化她的研究流程，她提及的功能例如“同时显示数据库中的 OCR 文本和图像，能够非常轻松地下载文件，根据 OCR 的可靠性排列检索结果，将文件组织到内容集中的过程非常迅速，清理 OCR 文本，以及使用多种不同的工具进行分析。”

“我对能够用来（几乎是立刻地）分析在 ECCO 中获得的 OCR 文本的工具感到惊讶且印象深刻。用户界面（UI）非常简洁且清晰，很高兴看到 ECCO 与实验室平台的整合毫无痕迹。”





内容集，列出了作者以及他们出现了 Atom（微量）一词的作品的数量

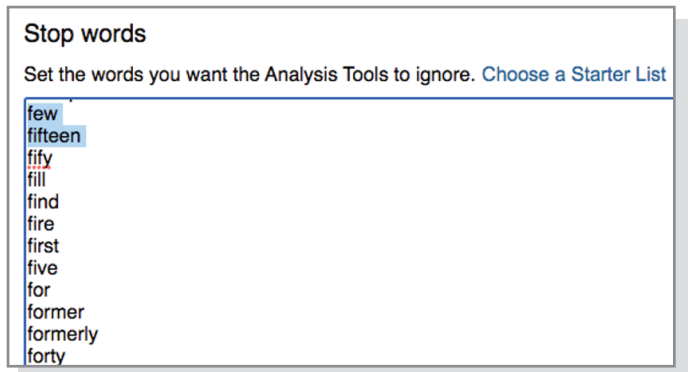
通过定制内容集克服难题

实验室引导用户分三个步骤完成一套直观的研究流程：构建（Build）、清理（Clean）和分析（Analyze）。在总结她构建内容集的方法时，Weston 说，“我原本只是采用精读文献的方法，以非常精确的方式讨论数量词（few [几个]、many [许多]、multitude [众多]、infinite [无穷尽]、nobody [没有人]、one [一个]等）。但转为使用实验室提供的大型数字工具后，我才意识到在如此庞大数量的文献中追踪这些数量词时，其中的一些词是非常模糊不清的。例如 myriads（无数）、hundred（百）和 thousands（千）这样的词可能易于追踪，但像 half（一半）、infinite、few、nobody 和 one 这样的词几乎无法追踪。”

为了帮助克服这一难题，Weston 说，“我最终不得不试验多种方法，调整研究项目，排除某些词，集中精力于更具描绘性的词语，以此得到更准确的结果。”

在优化检索标准后，Weston 在构建内容集时有了新的发现。“我认为最吸引我的一点是查看内容集时看到实验室列出了内容集中的作者名单，并标注了内容集中每位作者作品的数量。因此，仅通过查看他们的作品在特定内容集中出现的次数，我就可以了解到一位作者对某个数量词的关注程度，” Weston 说。

“我很感谢这个奖学金带给我的机会，训练我使用如此多不同的工具和如此多新的技能，为我提供了很多解决问题和找到问题的新经验。”

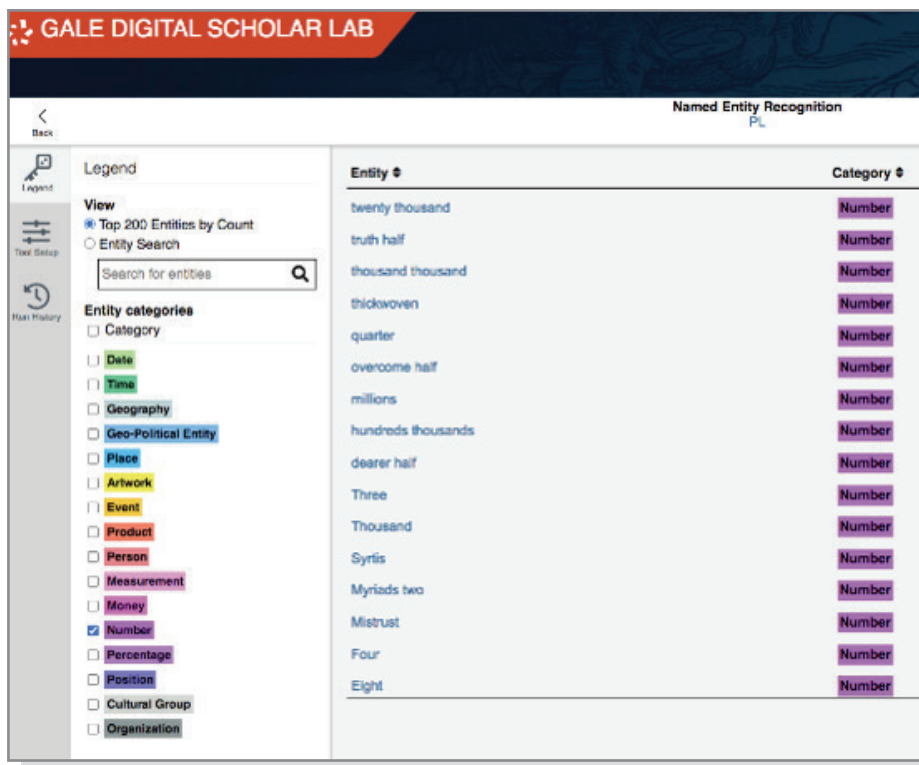


预备分析用数据

用户在实验室中构建他们的数据集后，他们可以定制自己的清理参数，预备用作分析的数据。

“‘清理’（Clean）工具对我来说是一项了不起的资源，” Weston 说。“能够立刻去掉标点符号、数字等，绝对是意义重大。在我之前的一个数字人文研究项目中，我不知道能够对 OCR 文本运行清理配置，花费了很多的时间细致地整理所有文本并自己进行清理，因此我知道手工完成这项工作需要多少时间。实验室的清理工具为我的研究项目节省了大量的时间，我真的震惊于这个过程。”





命名实体识别，显示数字类型的分析结果

“在奖学金期间通过实验室使用过不同的分析工具之后，我感觉我充满了全新的活力回到我正在进行的研究项目中，急切地想要使用我在奖学金期间熟练掌握的新工具。”

利用分析工具发现新见解

在分享 Gale-ASECS 奖学金为她的研究带来的好处时，Weston 指出，“这项奖学金最宝贵的地方在于我能使用的分析工具种类很多。在奖学金刚开始时我并不完全知道哪些分析工具或方法会对我最有帮助……我很感谢能有所有这些工具任我使用，最终我找到了对我最有价值的方法。”

Weston 解释说，“命名实体识别（Named Entity Recognition）最终是对我最有帮助的工具之一。这出乎我的意料，因为我曾认为命名实体识别主要是用来寻找人名和地名的。然而，我很高兴地看到我也可以用它来分类‘度量’（Measurement）和‘数字’（Number），它们最终为我的研究项目产生了最丰硕的成果。”

发现定量数据以支持理论

实验室为 Weston 提供了她所需的工具，证实了她在成为奖学金学者之前无法证实的假设。“我发现的最重要的证据是我在检索 heaps（堆）、piles（叠）和 masses（大众）时，得到了非常大量的检索结果是有关奴隶制和废奴的作品，” Weston 说。“我之前曾研究过废奴运动或有关奴隶制的作品中描绘的堆的意象，但我的文献样本数量很小（Equiano、Clarkson 等人的作品），我从未意识到它比我想到的还更加普遍存在。不使用实验室我不可能会偶然发现这个信息。”

实验室让 Weston 能够发现更多定量数据以支持她的理论。“现在我可以自信地断言，很多很多作家都普遍在废奴文学作品中使用了相同的意象和对这些特定词语的青睐，”她说。

Weston 给出了一个具体的例子，“检索结果中的一条来自一份报告《经多位商人分别证实的非洲贸易的改

善》(The Improvement of the African Trade Farther Demonstrated by Separate Traders, 1708 年)……这次检索为我提供了一个例子，这部十八世纪早期关于奴隶制的作品没有充斥着堆、叠、大众、众多等这样的意象，与我所习惯的十八世纪晚期作品完全不同。”

在意识到这一点后，Weston 从一条新的路径打开了她的研究。“我开始有意扩大我的检索范围，以十八世纪早期创作的作品为目标……我能够找到许多这样的例子：作品谈及奴隶制但避免使用偏向负面的数量词。这帮助我验证了最初的评价/猜想——在讨论奴隶贸易时堆和叠的意象、被折磨的大众、对种类和数量之多的焦虑感，是十八世纪末特有的一种情感，出现于废奴时期，是人们对数量和数字的一种新态度：对受奴隶贸易影响之巨大人口规模的测算。

The Improvement of the African Trade farther Demonstrated by Separate Traders, in Answer to a Scurrilous Paper, called, The Falsties of Private Traders Discovered.

THE Computation of 25000 Negroes per Ann. carried to the Plantations by Private Traders, was grounded on the Value of their Exports, as well as the Number of Ships Employ'd in carrying them; for the Time of Nine Years in an Averige, and not on the Six Years of War, when the Spanish Trade was loft, which took off Six or 7000 Negroes per Ann. and appears well-grounded from Political Arithmetick, with just Allowances, Confirmed by many Examples, of Ships Voyages, and Demonstration fo plain, that the Separate Traders need not retire to Sophistry, Falfe Arguments, and Falfe Quotations, as the African Company have done, and appears by their Accounts of the Private Traders Exports, given into the Right Honourable the Council of Trade, as well as taken out of their own Books, and compared with their Printed Account, fo full of Unjust Reflections, Published the 25th Instant, they differ near 40000*l.* and conceal the vast Exports of the Separate Traders in the Preceding Years of Peace, which stand thus, viz.

Account of 10 per Cent. Duty receiv'd of Separate Traders to Africa, between 1699 and 1707.				Account of Six last Years Exports, as quoted in the Companies Scurrilous Paper.			
Anno	l.	s.	d.	l.	s.	d.	
In Peace, 1699	5961	7	11				
1700	11584	10	7				
1701	11353	10	9				
1702	5363	8	3				37875 18 6
1703	4413	6	2				44115 12 3
In War, 1704	3997	3	0				26527 7 11
1705	3568	19	0				30651 7 6
1706	3127	7	10				32144 19 6
1707	3789	4	4				31986 16 8
	53158	17	10				203302 2 4

Whereof in Three Years of Peace 28899*l.* 8*s.* 3*d.*
in Six Years of War 24259*l.* 8*s.* 7*d.*

Now the 10 per Cent. Duty paid by Separate Traders being a Guide to know the Value of their Exports, and deducting a small Matter of that Duty, being

《经多位商人分别证实的非洲贸易的改善，对一份名为〈揭露私人商人谎言〉的文章的回复》，1708 年，ECCO。

思考实验室未来的应用潜力

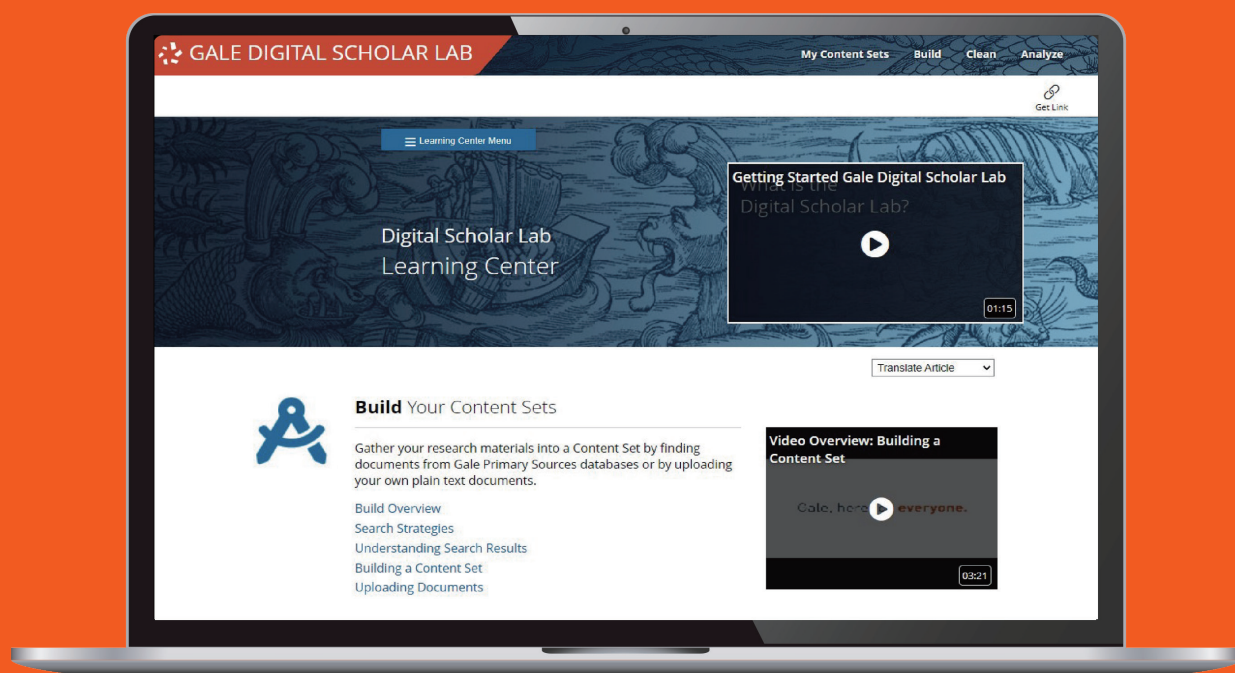
在接受实验室培训并在 Gale-ASECS 奖学金期间使用实验室开展研究之后，Weston 认为她面前有很多利用数字人文工具开展研究的机会。”我现在具备了新的技能，我会继续在我的研究和教学中磨练和改进这些技能，”她说。

对于她现在的著作研究工作，Weston 指出，“我现在有大约 50 部新的文献资料可以进一步精读、研究并结合到我的专著中。只有使用实验室我才能够发现并找出这些文献……我可以扩展研究项目的视野，将之前没有在我视线内的新类型的文献包括在内。这个项目现在变得更加丰富，更加广博。”

因为看到了实验室在教学环境中的价值，Weston 补充说，“我也计划将这些工具带入到课堂中，我感觉很有信心开发一些课堂计划，教授学生如何使用工具并找出哪些工具最适合他们的研究目的和需求。”

“我的经历整体非常积极，我会根据研究项目和个人需求，推荐各种工具给我的同行研究者或学生，”Weston 说。“我可以看到所有六种不同的分析工具都对不同的研究项目很有帮助。文档聚类 (Document Clustering) 能够帮助到的同事和学生例如他们正在数据库中搜寻一些文献，这些文献与他们已经找出的、认为有用或有研究兴趣的文献相类似。主题建模 (Topic Modeling) 能够帮助到的研究者例如他们已经有了一套无序的数据，想要更好的理解这些数据，找出常常在数据集中同时出现的词语。现在我已经熟练掌握了所有这些工具，很期待向我未来的学生们做量身定制的推荐。”

“我总体上感觉已经做好了准备，使用这些工具进行教学，以及使用它们开展新的研究项目并带着新的见解和我已经使用自如的工具回到原有的研究项目中。”



查看更多有关 Gale 强大研究工具的详情

想要对 Weston 用来提升研究水平并培养数字人文技能的 Gale 资源有更多的了解？

发现 Gale 数字学术实验室带来的各种可能性：

gale.com/digital-scholar-lab

找到更多有关 ECCO 的信息：

gale.com/ECCO

想要更多了解美国十八世纪研究学会？

AMERICAN
SOCIETY FOR
18TH-CENTURY
STUDIES

发现更多有关 ASECS 的详情，包括学会的项目计划和活动等，请访问：asecs.org