

发 挖 数 字 人 文 的 潜 力

GALE DIGITAL SCHOLAR LAB

Gale数字学术实验室

平台更加先进、工作流程和团队协作更加流畅、学习中心内容更加充实，新手和资深研究者都可以利用Gale数字学术实验室开启创新研究之路，这套直观、基于云计算的分析工具将帮助研究者用数据讲述更具冲击力的视觉故事。

驾驭数字学术的力量



“Gale数字学术实验室为我们的读者提供了独一无二的机会，以他们可能从未想象到的方式使用一次文献，而不需要提前具备编程知识或预先清理数据。虽然实验室平台为研究带来了巨大的潜力，但我们也看到这是一项出色的教育投资，因为实验室平台让我们能够利用文本挖掘范例为学生提供实际的使用经验，学生们可以将这些经验用在他们的研究中或他们大学毕业后的工作中。”

—Katrina McAlpine, 澳大利亚悉尼大学
出版和数据服务部副主任



在图书馆员、研究者和学生反馈意见的帮助下，Gale改进了Gale数字学术实验室的效率和使用体验，帮助用户从历史研究中获取有价值的见解。实验室平台将图书馆定位为院校的数字人文（DH）枢纽，帮助专业水平各自不同的学生在进入职场前掌握二十一世纪的技能。

使用GALE数字学术实验室帮助你：

- **面向所有人推广数字学术。**不仅为学生，也为经验丰富的研究者消除进入数字人文领域的障碍。通过简单培训即可组建、清理和分析自定义数据集。此外，定期的平台功能升级和完全嵌入在平台内的学习中心（Learning Center），都让新手用户能够最大程度地从中受益。
- **加大人文专业技能培养范围，为就业做好准备。**学生熟练使用Gale数字学术实验室的过程中将培养起批判性思维和分析技能，二者对各行各业的工作都非常重要。
- **提高图书馆投资的价值。**投资管理状况良好的一个表现是，有效利用实验室平台及现有Gale原始档案（Gale Primary Source）数据库的图书馆往往对这些资源的使用量和熟悉程度更高。
- **成为数字人文的未来。**通过Gale数字学术实验室推动Gale原始档案数据库的使用量，院校将能够充分利用实验室不断拓展的文本分析能力。从数千万页Gale独有的一次文献内容中生成有意义的研究结果，这将创造一个前所未有的机会，生动地讲述出怎样利用图书馆的数字馆藏从事研究。

三个直观的步骤可视化数据

Gale数字学术实验室将复杂的检索、清理和可视化过程简化为三个明确的工作流程，充分利用六种文本分析工具，节省时间且提高研究成果。



步骤1：构建

- 组建一个文献语料库，提供可用于生成可视图表的原始文本数据。
- 在Gale数字学术实验室平台上检索图书馆的Gale原始档案馆藏，或上传任何本地保存的纯文本文件。
- 通过并排查看光学字符识别（OCR）文本（原始文本数据）和Gale原始档案中的文献原件扫描图像，快速确定数据的准确性。
- 选择加入到定制数据集中的文献，以备清理和分析——每个内容集最多10,000篇文献。
- 新手研究者也可以选择从学习中心导入预先构建好的数据集。



步骤2：清理

- 为预期将要使用的分析工具设定内容集的清理配置。
- 测试内容集中的一小部分文献，确定所选择的清理配置适用于预期的分析工具。
- 在工具设置选项中选择调试完成的清理配置。



步骤3：分析

- 构建和清理定制内容集后，利用文本挖掘和可视化工具分析数据。
- 平台收录六种常用的数据挖掘工具：N元语法、命名实体识别、主题建模、词性标注、情感分析和文档聚类。
- 每种工具的开发都征询了图书馆员和教师的意见，服务于数字人文学界公开透明的核心价值。

刚刚接触数字人文？

Gale数字学术实验室旨在支持所有人的数字人文研究。平台随处都可找到清晰的描述且持续改进升级。学习中心的帮助文档以及随时可取用的数据集，都为新手用户提供便利。

高级用户的定制选项

有文本和数据挖掘（TDM）经验的用户可以在Gale数字学术实验室中优化他们的研究：

- 利用更大范围的检索和无缝衔接的优化流程，快速构建和清理内容集——只需几分钟而不是几天的时间就可以完成内容准备。
- 自定义工具配置，探索和优化数据集。
- 下载清理后的数据集和元数据，用他们自己的程序（例如用Python或R语言编写的程序）或其他数字人文生态系统中的统计工具，在本地进行分析。

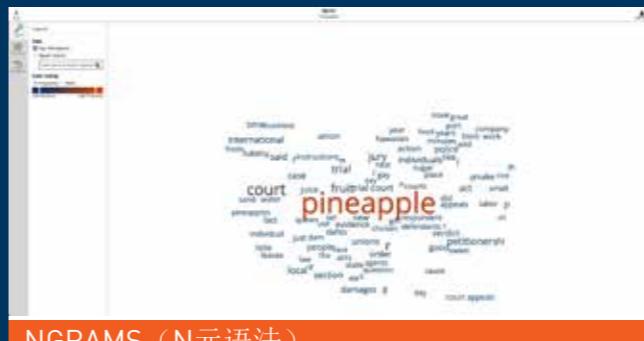
六种强大的数据分析工具

因为强大而易于使用，实验室平台上的六种分析工具提供无数种方式用研究数据讲述视觉故事。



DOCUMENT CLUSTERING（文档聚类）

使用统计学方法分析文档并将它们分组，以确定你的内容集中每篇文档之间的相似性。



NGRAMS（N元语法）

根据用户想要考察的词组长度（N）计数，并将内容集中词组出现的频率显示为“词云”。



SENTIMENT ANALYSIS（情感分析）

通过向每一个词语指定正值或负值并在整篇文献中平均这些数值，得到内容集中每篇文献的总体情感值。



NAMED ENTITY RECOGNITION（命名实体识别）

从文档中识别和提取专有名词和普通名词，将它们按实体类型，例如人物、组织、公司、地点等分组列出。



PARTS OF SPEECH（词性标注）

利用句法的自然语言处理，识别和标记不同的词性，为用户呈现句子是如何构建的。



TOPIC MODELING（主题建模）

帮助用户分析一大组无结构的文本，并将常常同时出现的词语分组为“主题”。

学生的研究之旅

从项目构思到取得成果

看看英国纽卡斯尔大学的英国文学本科课程是如何利用Gale原始档案并在Gale数字学术实验室的助力下探索数字人文研究方法、遥读和精读，以及定量分析结果和定性分析结果之间的关系。^{*}

构思

在这门课上，学生们被要求探讨十八世纪英国文学中边缘人群的故事——特别是被解放的奴隶的作品。



发现

学生们可以选择使用Gale原始档案的“十八世纪作品在线”中的两部专著，或他们自选的其他文献资料。



构建

学生们上传他们选好的文献资料到Gale数字学术实验室的自定义内容集中。



分析

学生们使用主题建模和N元语法工具，输出一个关键词列表，用于为这门课做出分析和评价。



清理

在学习中心教学视频的帮助下，学生们对他们的内容集应用“停止词”和其他限制条件，为分析做好准备。



“当然，尽管学习文学是主业，但我们也帮助他们为迈出校园后的生活做好准备。”

—Leanne Stokoe博士，纽卡斯尔大学英国文学、语言和语言学学院十八世纪文学讲师

使用GALE数字学术实验室的好处

节省时间：如若没有Gale数字学术实验室，文献资料将需要逐页进行解读。实验室的分析工具将这个传统上可能需要几天时间的过程压缩到几分钟。

培养宝贵的技能：数据显示，学生们开始在课堂之外主动使用Gale数字学术实验室完成他们的作业。

提高平台使用量：这个项目推动纽卡斯尔大学Gale原始档案数据库的访问次数增加了14%，与上一学年相比检索次数增长8%，在课程研究阶段，创建的内容集数量比上月增长了78%。大学计划将对实验室平台的使用纳入到这门课的成绩评定中，而且已经将实验室平台用于其他教学课程中。

认识我们的团队

这是Gale的研究和开发人员，他们从零开始搭建起Gale数字学术实验室并仍在不断改进升级这个平台。



Jess Ludwig

是Gale数字人文产品管理总监。她领导着Gale原始档案和Gale数字学术实验室的全球数字平台战略，强调创新性和可持续性，以及最为重要的用户需求。



Rebecca Bowden

是Gale数字学术实验室产品经理。她热衷于探索数字人文的世界，并且满怀热情地与我们的用户展开长期合作，开发和提升实验室平台。



Chris Houghton

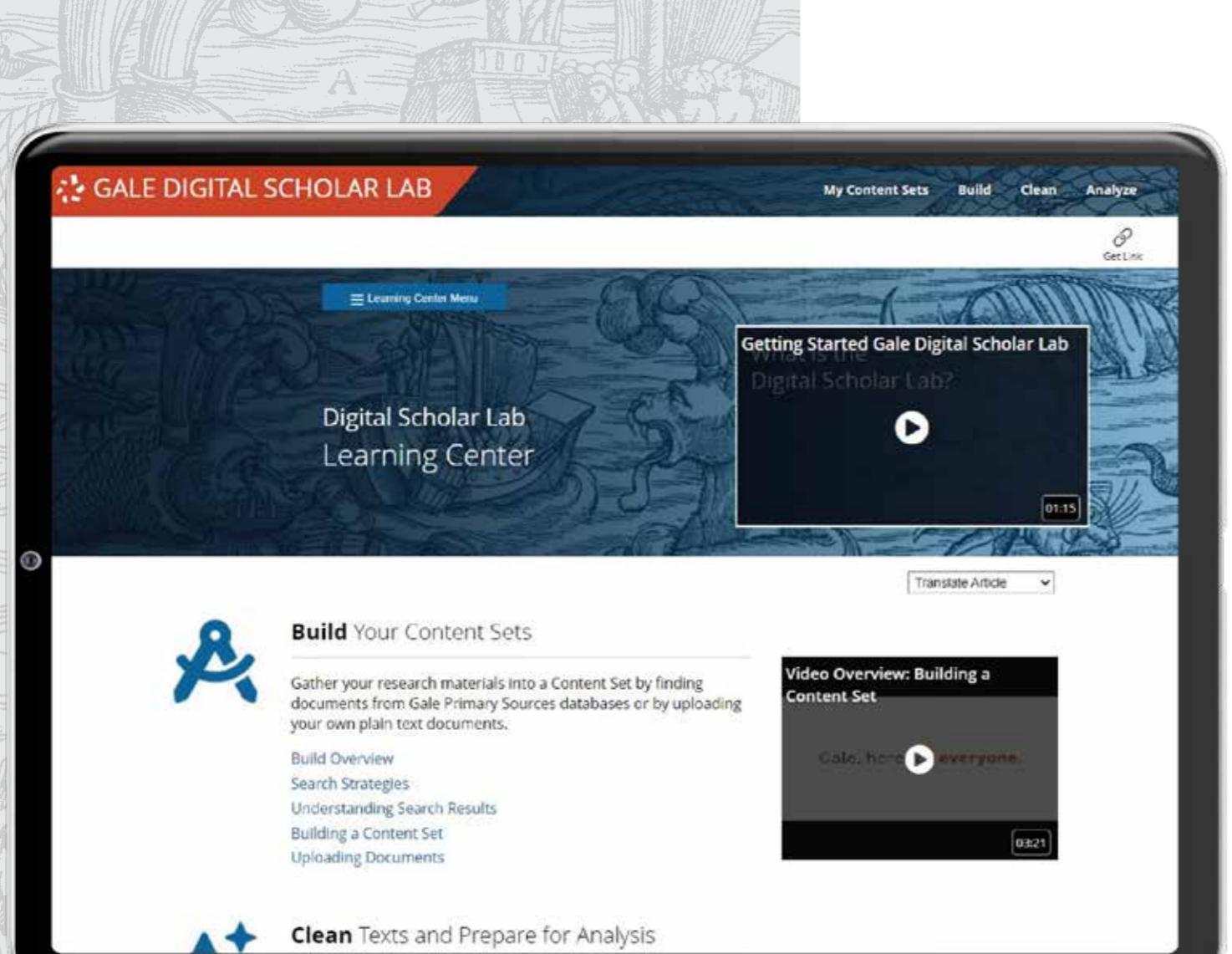
是Gale学术合作主管。他与全球数字人文学界保持紧密联系，通过与学者的协作和咨询开展教学并支持研究项目。



Dr. Sarah Ketchley

是Gale数字人文高级专家。她是古埃及学博士，并在华盛顿大学近东语言和文明系任教，开设数字人文的入门课程和研究生课程。

让所有人都能从事数字学术研究



学习中心（Learning Center）内置在实验室平台中，令Gale数字学术实验室成为了一个理想的环境，在此能够培养对数字人文研究方法的深入理解，并随时可以开展有意义且可复制的研究。

图书馆员和讲师可以利用学习中心指引新手研究者，帮助他们在工具和数字人文数据的各种可能性中找到方向，为他们未来的研究之路培养批判性思考技能。

利用学习中心开展数字人文研究方法的教学

- **全面整合：**学习中心完全嵌入到Gale数字学术实验室的构建、清理和分析工作流程的每一个步骤中，营造出一种氛围，既是专为刚刚接触数字人文的用户而开发，同时也为更有经验的用户提供一系列直观的用户体验。
- **一站式学习：**提供演示操作步骤的教学视频、术语表、常见问题解答、项目构思指南以及图书馆馆藏数字档案清单等，均为精心制作。学习中心里完整研究项目范例的数量将持续增加，基础内容集专为消除平台使用的障碍，加速研究经验的积累。
- **各种类型的研究者：**学习中心不仅是学生们的重要工具，也增强了图书馆员和教师的教学能力，为提高整个院校的数字人文技能和数据素养提供了一套一站式资源，将Gale数字学术实验室与课堂教学联系在一起。
- **学生顺利掌握数字人文技能：**在学习中心的教学支持下，学生研究者能够迈出重要的一步，逐渐掌握可转化到未来工作中的数字素养和统计分析技能。如果为就业做好准备是目标，那么在Gale数字学术实验室的学习中心里学习研究技能就是通向这个目标的道路。

在GALE数字学术实验室中创新

集体项目，更快得到结果，更丰富的见解

Gale不断升级Gale数字学术实验室平台，让用户能够提出更多问题，在Gale原始档案的内容中找到更多相关检索结果——并且在集体研究项目中协作。

- 拓展检索功能：**检索选项与图书馆现有的Gale原始档案数据库相一致，包括OCR置信度筛选条件，有助于为分析步骤创建高质量的内容集，快捷链接便于从档案库快速进入分析工具。
- 区分工作流程：**优化后的构建、清理、分析框架版式强化了用户与可视化图表的互动，能够提升对主题的探索。
- 快速查看并评价：**新的分析工具面板让研究者能够对他们的工作进度一览无余。



平台协作

Gale数字学术实验室最新上线了群组（Group）功能。在专用的工作区内实现项目协作，同一院校内的多名用户均可在此添加内容集、注释和分析结果。

案例

Gale数字学术实验室用于本科生教学：数字人文如何能够推动数据素养并发掘重要史实

清除数字人文的入门障碍

Catherine A. Nichols博士是芝加哥洛约拉大学的文化和人类学及博物馆研究高级讲师、梅·韦伯人种学研究典藏主管。作为一名社会科学工作者，她受到鼓舞的一点是她的学生将能够使用计算方法处理数据。虽然她认为传授一名学生编程技巧来完成这项工作是可能的，但她知道如果是在人文专业本科生的大课上，这么做则是非常困难的。

Gale数字学术实验室的出现帮助Nichols博士消除了课堂教学的障碍，为她所有的学生带来了文本和数据挖掘的能力，不再需要编程知识。

与Gale数字人文专家Sarah Ketchley合作，Nichols博士将Gale数字学术实验室整合在她的教学中。她最先在她研究生的人类学课堂上介绍了实验室和Gale原始档案，指导学生创建内容集，分析的主题例如西方传教士收集的博物馆藏品以及数字时代的知识生产等。

支持学生在GALE数字学术实验室中顺利开展学习

2021-2022学年，Gale数字学术实验室的优势变得更加明显了，在这个学年Nichols博士向她的100名人类学本科学生介绍了实验室平台。这一项目式教学课堂的教学目标包括建立全球力量对文化影响随时间改变的理解。

为了达到课程目标，Nichols博士要求她的学生在Gale数字学术实验室中创建内容集，并对内容集运行三个分析工具——N元语法、情感分析和主题建模。她编写了一份指导文档，向学生们介绍实验室及其定量和定性研究方法。随后学生们按照工作文档模板中的说明和问题，使用这些资源生成可视化分析结果，下载并用在他们的作业中，论述分析结果，例如为什么结果是正面的或负面的。

见证影响力

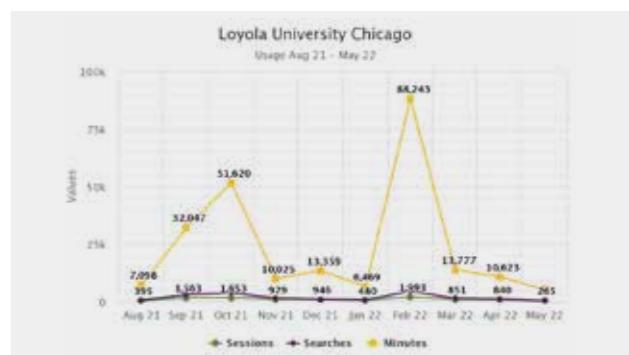
在本科课程开设期间，该校的实验室平台使用量大幅攀升。学生们显然在Gale数字学术实验室中花费了大量的时间使用图书馆的多个馆藏，在2022年2月达到了88,000分钟之多。这个月内Gale原始档案的下载量也比上一年2月提高了30倍，平台访问时间翻了一番。仅在一个月内，学生们就将各个Gale原始档案数据库中的222,000多篇文献加入到内容集中。

“随时可使用的数字人文数据以及分析这些数据的工具，令我能够让我的学生们真正参与到研究中。”

—Dr. Catherine Nichols
芝加哥洛约拉大学
文化和人类学及博物馆研究
高级讲师

在课程过程中，Nichols博士和她的学生们创建了近1,225个大小各异的内容集，内容来自于多个Gale原始档案数据库，每名学生平均创建了17个内容集。主题范围从移民、医疗保健全球化和气候变化，到青铜器时代的文明及十字军东征对东、西方的影响。

洛约拉大学的用户也运行了数百次N元语法、情感分析和主题建模分析工具。这些数字说明他们使用了Gale数字学术实验室中提出的迭代研究过程。



2021年8月至2022年5月洛约拉大学对Gale数字学术实验室的访问次数、检索量和访问时长（黄线）。



联系我们

圣智学习（北京）教育科技有限公司

北京市海淀区科学院南路2号融科资讯中心C座南楼707室

邮编：100190

电话：+86 10 8343 5000

传真：+86 10 8286 2089

邮箱：GaleChina@cengage.com

更多有关Gale数字学术实验室的详情，请访问：

gale.com/digital-scholar-lab

欢迎关注Gale官方微信号：

Gale 数字学术

