

Gale数字学术实验室

# 发现各种可能性

透过新的视角探索一次文献

构建



清理



分析



# 来自Gale团队的 见解

## 我们从何而来、我们去往何处

### Seth Cayley, Gale原始档案项目副总裁



Gale多年来一直站在前沿地带，支持数字领域的人文研究。在过去的十年间，我们与客户合作，尝试理解我们的档案资源在文本和数据挖掘研究项目中的应用。很快我们就意识到，用户在试图分析大型档案文献集时常常面临着一些挑战。为了解决这些难题，我们在全球范围内招募专家顾问，开发了一个将常用开源工具与Gale领先档案数据整合一体的平台。而且，这是一个既能用于研究又能用于教学的平台。

这就是“Gale数字学术实验室”。这一全新的平台汇集了多个学科领域内600多年的一次文献资料和一整套工具，令学生和研究者能够通过学习和使用其中便捷的开源工具对内容提出新的问题。

随着新的特色和功能不断加入到这一平台中，Gale热切地期待与新老合作伙伴一同努力，更好理解校园内的研究、教学和学习需求。最为重要的是，我们依赖图书馆合作伙伴的反馈意见，据此进一步开发实验室平台，让它能够更加凸显图书馆在支持各个课程学术研究中发挥作用及其所拥有资源的价值。

## 探索发现的新语言

### Chris Houghton, Gale国际部数字学术主管



实验室平台在发布后的前18个月中经历了长足的进步，而进步的动力来自于客户的建议、指导和检验——折射出真实的二十一世纪学术研究过程。

除了与客户保持互动，我们还使用实验室平台的应用数据，不断修正它的发展方向。例如我们增加了OCR下载的数量——在实验室平台发布之初，研究者能够一次性下载的文档上限是每次1,000篇，但通过分析使用量我们发现，自发布起在实验室平台上创建的内容集中，有5.5%超过了5,000篇文档，因此新的下载上限设定为了这一数量。

我们也听到很多客户很感兴趣在实验室平台上传自己的内容，然后进行分析，因此我们新近增加了这一功能。我们的目标是为每一个人——无论他掌握的技能如何，提供一整套工具和一种语言，不仅用于研究数字人文，也可以用于传统学科，实现数字研究概念的整合，为探索和发现提供新的视角。

# 实验室平台上的学习

Gale数字学术实验室非常简便省时。如果你用传统方法，研究会非常费时费力，但如果你使用实验室平台上的工具，研究会变得更加简便。”

— 许明杰博士，复旦大学英国史讲师

## 帮助所有人走向数字学术

随着“Gale数字学术实验室”的不断进步，在实验室平台上出现了很多学习活动，因为这一平台能够辅助对整个研究过程的教学和指导。通过“学习中心”内的资料，实验室平台助力图书馆员、导师和学生们，改进他们的教学、学习和学术研究过程。“学习中心”为用户提供了他们理解和使用实验室平台上海量信息和常用技能所需的工具与技术支持。它包含由我们的数字人文专家主讲的实际操作演示视频、研究项目范例、术语表、常见问题等。

这些延伸至实验室平台本身以外的教学资料让用户具备与研究项目紧密结合的背景知识，培养思考研究问题和现象以及解读数据和发现所必需的核心思辨能力。通过支持用户“超越”实验室平台，我们与大学院校一同，努力推动数字人文等新兴学科领域的茁壮发展。

## 挑战蕴藏机遇

### 1

难于访问和收集权威内容可能导致一些学者花费数年的时间收集资料、数个月的时间准备内容集。

▶ 大大减少收集高质量内容集所需的时间。

### 2

图书馆遭遇内容存储以及为研究者提供内容访问的难题。

▶ 揭示海量云存储的档案资料。

### 3

无尽的可能性和大量的各种工具让新手研究者很难应对数字人文数据。

▶ 为用户提供更易于使用、更直观的工具。

## 营造简化的研究体验

“Gale数字学术实验室”通过三个直观的步骤引导用户，简化了复杂的工作流程，改进了研究结果。

### 步骤1 构建

- 研究的第一步，用户需要构建一个文献语料库，提供原始文本数据以生成可视化图形。
- 用户检索他们所在图书馆的“Gale原始档案”（Gale Primary Sources）馆藏，选择文献加入到他们的自定义内容集中。
- 并排查看OCR文本（原始文本数据）和文献原本的扫描图像，帮助用户快速确定OCR文本的准确性。
- 用户可在实验室平台上传本地文本文件并添加元数据，与来自“Gale原始档案”的文献资料组成混合内容集。

### 步骤2 清理

- 作为“Gale数字学术实验室”的一部分，用户可以创建多个清理配置，根据分析的需要设定如何清理内容集。
- 测试清理配置是用户确定是否已准备就绪可以开始分析的一个重要步骤。在“Gale数字学术实验室”中，用户可以选择一个内容集开始测试他们的清理配置，随后内容集中的前10篇文档将以此清理配置清理。然后，用户可以下载这些文档的原始版本和清理后版本，确定文档清理的程度是否符合他们的需要。
- 用户对他们的清理配置感到满意后，就可以将其应用于他们的分析过程。用户可以在分析工具设置中选择他们想要使用的清理配置。

### 步骤3 分析

- 在选择文献并创建自定义内容集之后，用户就可以使用文本挖掘和可视化工具分析和考察数据了。
- 平台包含六种常用的文本挖掘方法：N元语法（Ngrams）、命名实体识别（Named Entity Recognition）、主题建模（Topic Modeling）、词性标注（Parts of Speech）、情感分析（Sentiment Analysis）和聚类（Clustering）。
- 清晰的描述和“学习中心”中的帮助资料为新入门的数字人文研究者提供支持。资深研究者则可以自定义工具设置，生成新的分析结果。

当我们开始考虑采购‘Gale数字学术实验室’时，我们大学正在筹备开设一门数字研究的辅修课。为了支持这门课程，我们必须为学生们提供对在线馆藏资源的访问，这些资源相当于一个数字沙箱：学生们能够在其中进行数字化的实验、研究和操作。实验室平台令我们得偿所愿，此外它还提供了一整套工具，随时都可以在这个几乎无障碍的环境中使用。”

— Hillary Richardson，美国密西西比女子大学本科生研究协调员及信息素养图书馆员

## 实验室平台 工作流程

**我们的想法：**  
**Lindsey Gervais 博士，**  
 圣智旗下Gale公司数字学习经理



Gale因助力学习而享誉全球，为图书馆员、教师和学生提供高质量且极具深度和广度的内容以及能够增进学习机会的工具。内容与有效的学习工具之间的这种搭配启发了我在数字教学方面的工作，为用户提供他们需要的教学工具，帮助他们在院校中达成积极的工作成果。随着最佳的教学方法示范不断被注入到‘Gale数字学术实验室’的工作流程和‘学习中心’中，我们为用户配备了理解和使用知识和技能所需的工具和框架。从实验室平台培养而来的技能不仅对学者取得成功至关重要，对我们这个不断发展和创新的世界中的每一位数字公民而言也极为重要。”

### 构建 清理 分析



**我们的想法：**  
**Maggie Waligora，**  
 圣智旗下Gale公司产品经理



在‘Gale数字学术实验室’开发的过程中，我们与数字人文领域的图书馆员和学者积极合作，寻求他们的指导。这些合作关系伴随我们走过了实验室平台的初创版、测试版和成品版，并且始终是平台长期发展战略的驱动力。我们与Gale的用户体验团队合作，与80多位用户进行了设计和功能概念的测试。这些步骤帮助我们对实验室平台的所有构成要素做出重要决策。这些对话最终促成了100多项功能和改进，都与我们在过去18个月中收集的用户反馈直接相关。随着我们不断提升‘Gale数字学术实验室’的整体用户体验，我们计划加强与新老用户的合作，确保我们能够帮助他们实现他们的核心研究目标。”

# 听听您的同行怎么说

## 阿德莱德大学如何推动院系变革及开展数字人文课堂教学

无论是在校园中或是网上的形象展示，阿德莱德大学（University of Adelaide）都透露出成长、发展和进步的强烈愿望。这一雄心壮志映射在结果中，阿德莱德大学在过去的六年中在全球大学排行中不断前进。对于人文系主任 Jennifer Clark 教授而言，“这意味着大量的全新课程，意味着各种应对措施、各种教学风格……各种探讨人文学科的方式。”正如文学院联络员 Helen Attar 所说，“‘Gale数字学术实验室’未来将会在其中发挥重要的作用。”

媒体和数字人文讲师 Aaron Humphrey 博士解释道，“阿德莱德大学是澳大利亚很少的几所你可以主修或辅修数字人文专业的大学之一。我们期待加强这个方面，为学生介绍更多数字人文中使用的工具和技术。”对于Humphrey博士而言，数字人文的优势在于它“让学生能够很轻松地开展前人从未涉足过的研究。”Clark教授认为这一优势对于实现院系和大学雄心勃勃的目标至关重要：“最有意思的历史研究是提出从未被提出的问题，或探讨从未被探讨过的文献资料。”

大学旨在为学生提供新的发现路径，而“Gale数字学术实验室”则是其核心所在。作为一名数字人文教师，Humphrey博士对于实验室平台兴奋不已。“我们用来探索的工具无比强大。在教室里，‘Gale数字学术实验室’平台的优势是让学生在快速构建一个内容集、发起讨论、查看结果，然后优化分析过程。”

“最有意思的历史研究是提出从未被提出的问题，或探讨从未被探讨过的文献资料。”

— Jennifer Clark, 澳大利亚阿德莱德大学教授

位于格林斯博罗的北卡罗莱纳大学很荣幸地领导着一个由IMLS（博物馆和图书馆服务研究院）资助的多州协作项目，名为“卡罗莱纳州数字图书馆网络”（CDLN）。在这一项目中，我们评估了几项共享数字图书馆设施，从最引人注目的新产品‘Gale数字学术实验室’开始。这一实验室平台提供了惊人数量的文本规律分析工具，可用于对历史文献进行学术探讨。我们的项目在未来两年将会发起一系列的竞赛，由每所协作大学中的研究者参加，获胜的研究项目将会获得现金奖励。我们的教职工对实验室平台这样的数字人文工具带来的可能性产生了浓厚的兴趣，不仅仅是为了他们自己现有的研究计划，而且因为这些项目将为本科毕业生提供的下一代学术训练。我们最喜爱实验室平台的一点是，它将先进的文本规律分析工具整合到便捷的平台界面中，几乎各个学科领域的学者都能够快速学会如何使用。”

— Martin Halbert, 美国北卡罗莱纳大学格林斯博罗分校教授

最为关键的是能够随时整合到研究者的工作流程中。Humphrey博士又说到，“我们的教学时间是10周。我没有足够的时间教会学生们用命令行编程。要让他们学会在Python中进行主题建模，课堂时间是远远不够的。我青睐‘Gale数字学术实验室’就是因为它令这类分析变得非常简便，让我们能够快速获得稳定且可靠的结果，用于在教学环节中深入剖析。”

Clark教授则发现了它的战略优势。“我们想要通过数字人文提供将科学世界和艺术与人文世界连接在一起的机会，”她解释道。对于院系来说的另一个优势是能够吸引学生。“我们可以告诉他们，‘你仍在做艺术与人文学科的研究，但你也能够获得[技术]技能。’”Attar附议了这一观点。“数字科学是增长最快的工作领域。任何具备数据管理、数据操作、数据可视化或数据辩论技能的人都将会是受欢迎的人才”，她说。Humphrey博士也认可它对学生的价值：“数字技术中最为至关重要的一些工作就是找出学科在哪里交叉，我们如何能够培养出既具备技术素养又理解人文情怀的学生……各种可能的机遇几乎超出我们的想象。”

‘Gale数字学术实验室’为我们的研究者提供了独一无二的机会，能够以未曾想象过的方式探索一次文献，而无需预先具备编程或数据清理的知识。虽然实验室平台为研究带来巨大的契机，我们也看到它是对教育的绝佳投资，因为这一平台让我们能够为学生们提供可动手实际操作的文本挖掘模型，这些模型学生们可以在他们的研究中使用，也可以在大学毕业后的工作中使用。”

— Katrina McAlpine, 澳大利亚悉尼大学出版与数据服务副主任



Gale 数字学术实验室

**LEARN  
MORE**

✉ [GaleChina@cengage.com](mailto:GaleChina@cengage.com)

🖥 [gale.com/discoverlab](http://gale.com/discoverlab)