GALE Digital Scholar Lab 数字学术实验室操作指南2024

EMPOWER™ CHANGE

GALE

GALE

Digital Scholar Lab

数字学术实验室

介绍

当开展分析时，发现、清理和组织数据以及对历史文本的自然语言处理（NLP）常常是一项令人望而却步的任务，特别是想要生成有意义的分析结果时。**Gale数字学术实验室（DSL）**扫除了这些障碍，让工作流程直观顺畅，研究者能够将更多的时间用于找到之前未被发现的数据、检验理论、分析结果和获取新的见解。

- 创建内容集

Gale数字学术实验室允许用户创建自定义内容集，每个内容集最多包含10,000篇文档。用户可以在图书馆现有的Gale原始档案馆藏中检索，顺畅地选择文档并加入到他们的自定义内容集中。

- 分析内容集

用户可以利用Gale数字学术实验室内置的文本分析和可视化工具分析和探讨数据。数字人文分析方法包括：命名实体识别（Named Entity Recognition）、主题建模（Topic Modelling）、词性标注（Parts of Speech）等。

- 管理和分享

用户的内容集保存在Gale数字学术实验室中，便于他们为长期项目管理他们的研究。用户可以发表他们的研究结果，完全保留所有的知识产权，也可随意分享他们的分析结果。

GALE

Interface 登录界面

Get Inspired With Gale Research Showcase

Explore text and data mining projects created in Gale Digital Scholar Lab at Gale Research Showcase. Then submit your own project for inclusion.

Build Your Content Sets

- Use Gale Primary Sources archives
- Upload and use your own text files
- Download content sets for use elsewhere

Learn more about the Build step »

Clean Texts for Computational Analysis

- Apply stop words to your analysis
- Use flexible options to target specific character removal
- Reuse configurations across Content Sets

Learn more about the Clean step »

Analyze Content in Powerful New Ways

- Visualize data from up to 10,000 documents at a time
- Explore individual documents overlaid with analysis data
- Download the raw data and your visualizations

Learn more about the Analyze step »

数字人文研究指导/学习中心

Build 创建内容集

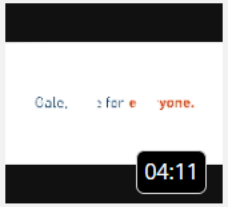
GALE DIGITAL SCHOLAR LAB

My Research | Build | Clean | Analyze

Add Note | Search History

Build

Build Content Sets by finding documents in your institution's Gale Primary Sources holdings or by uploading your own documents.



Video: Building a Content Set

Additional topics:

- Search Strategies
- Understanding Search Results
- View All Build Help »

Search

Search by keyword

Advanced Search

Refine your search by publication date, source ore

View all limiters in Advanced Search »

Logical Operator Examples

AND: feminism AND gloria steinem

OR: comic book OR cartoon OR graphic novel

NOT: war NOT world war II

COMBINATION: cloning AND stem cells NOT sheep

检索输入框/高级检索

通过检索，添加Gale历史一次文献资源

Upload

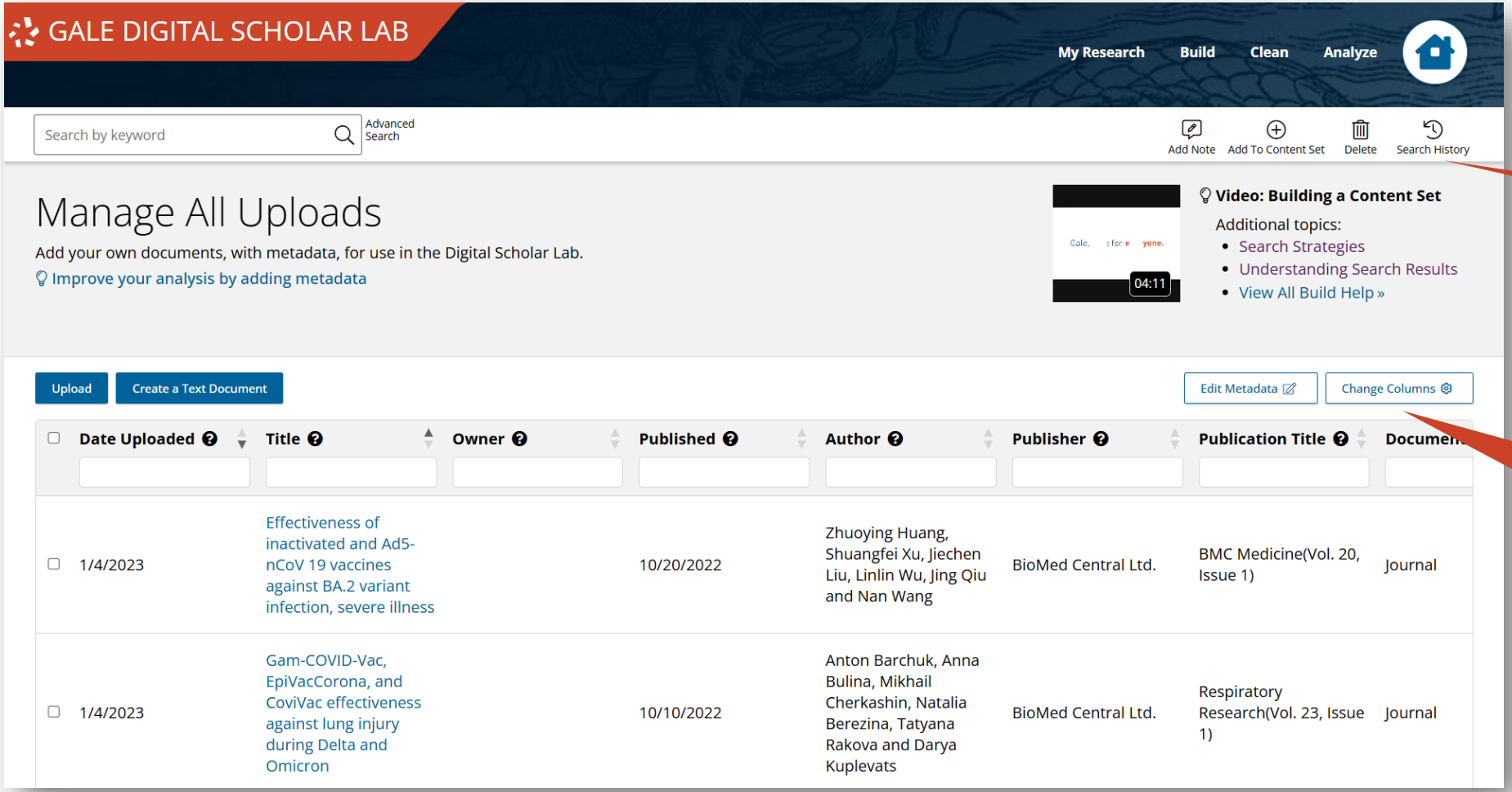
通过上传txt格式文本或手动编辑创建内容集
添加元数据便于文献管理

Drag and Drop .txt or .csv Files here or Brov

Get Upload Template | Create a Text Document | Manage All Uploads

10 MB at a time

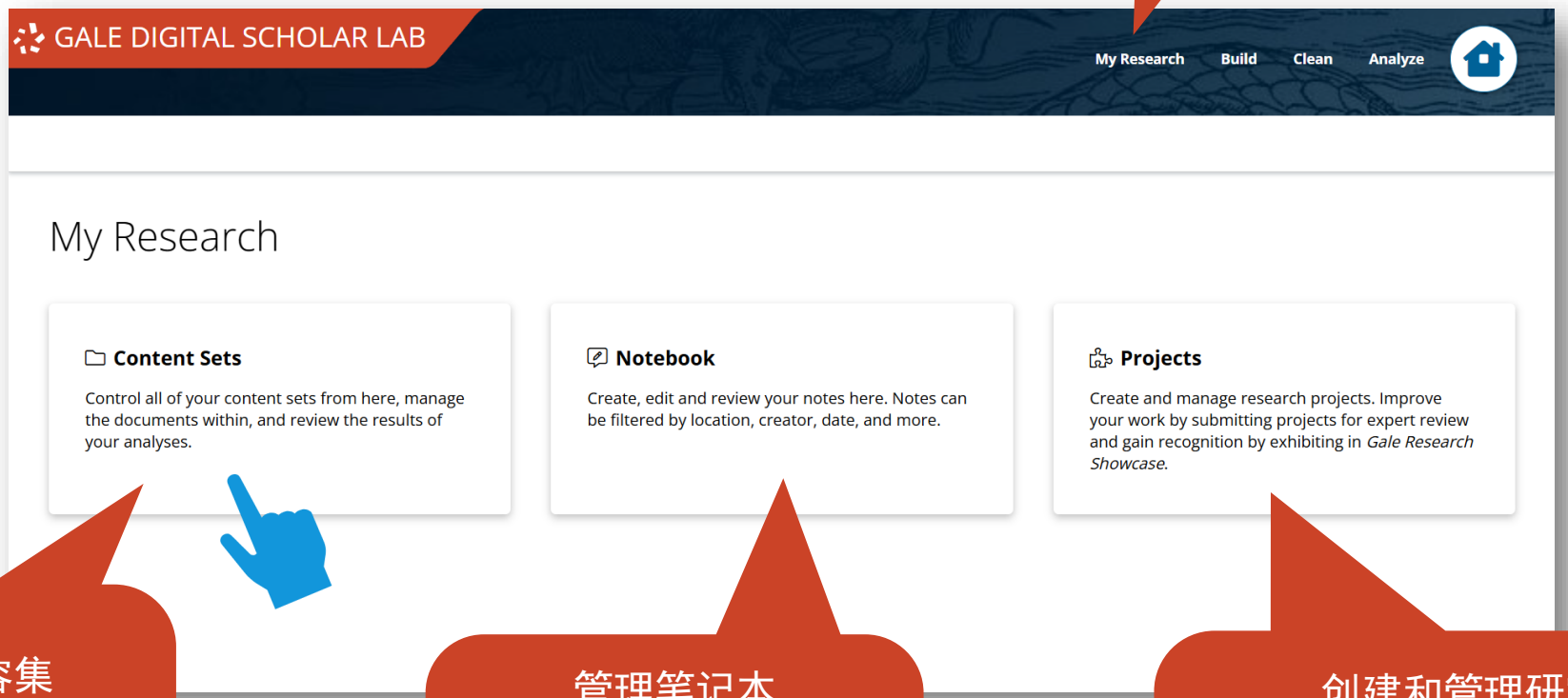
Manage Uploads 管理上传文献



管理上传文献

管理显示项目与元数据切换排序等

My Research 我的研究



管理内容集

从这里控制所有内容集，管理其中的文档，并查看分析结果。

管理笔记本

在这里创建、编辑和查看您的笔记。笔记可按地点、创建者、日期等进行筛选。

创建和管理研究项目

创建和管理研究项目。通过提交项目供专家评审来改进您的工作，并通过在 Gale 研究展板 上展示来获得认可。

Content Sets 内容集

GALE DIGITAL SCHOLAR LAB

My Research | Build | Clean | Analyze

Add Note | All Projects | New Content Set | New Folder | Move | Merge | Duplicate | Delete

My Content Sets

Control all of your content sets from here, manage the documents within, and review the results of your analyses.

Add Content | Manage All Uploads

添加笔记/研究项目/创建内容集/创建文件夹/移动内容集/合并内容集/复制内容集

Video: What is a Content Set?

Additional topics:

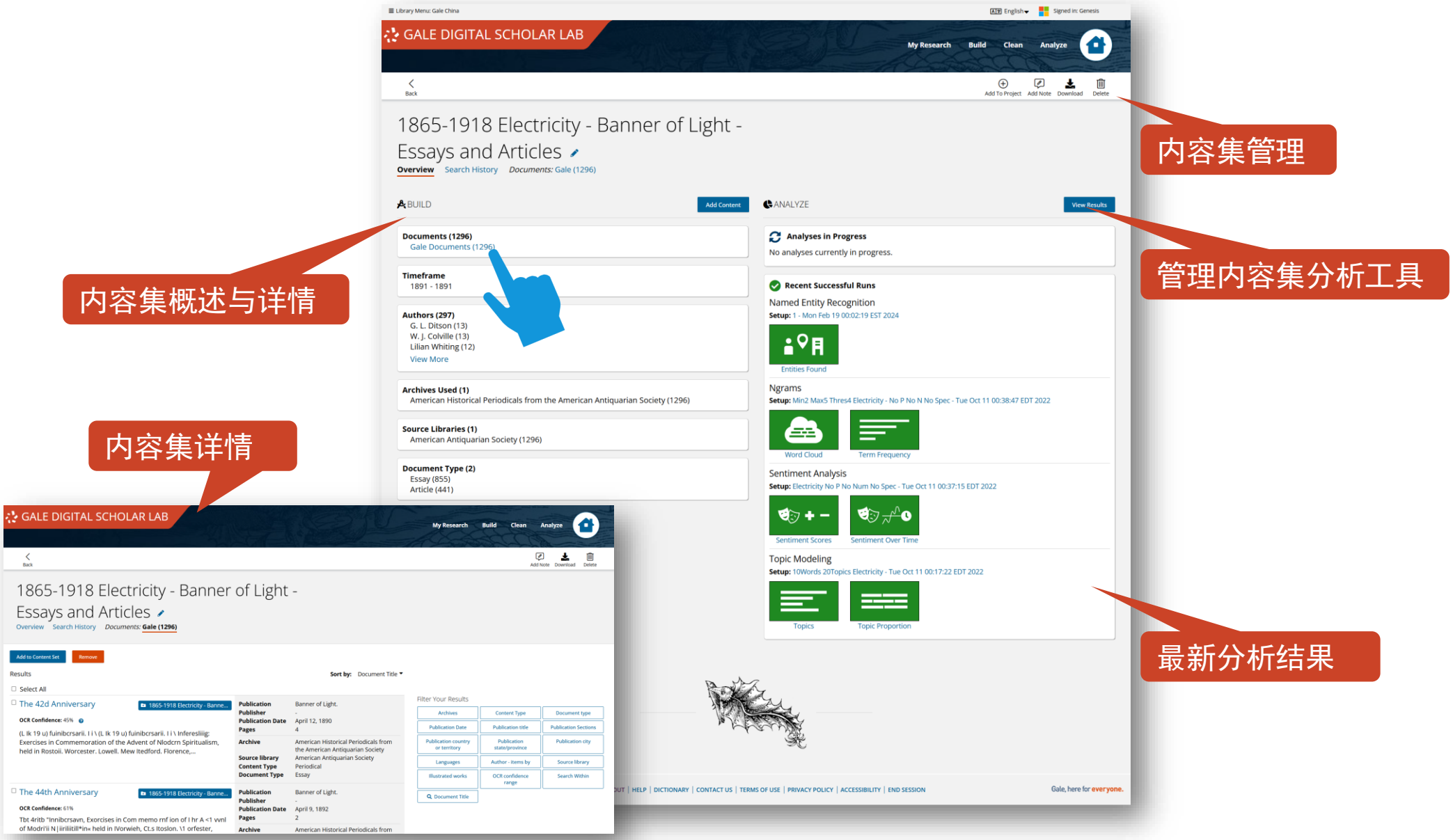
- Downloading Content Sets
- Sample Projects
- View All Content Set Help »

Type	Name	Build	Analyze
	AF	Content Sets: 13	
	law	Content Sets: 4	
	1865-1918 Electricity - Banner of Light - Essays and Articles	Documents: 1,296	Tools: 4 Runs: 7
	war	Content Sets: 5	

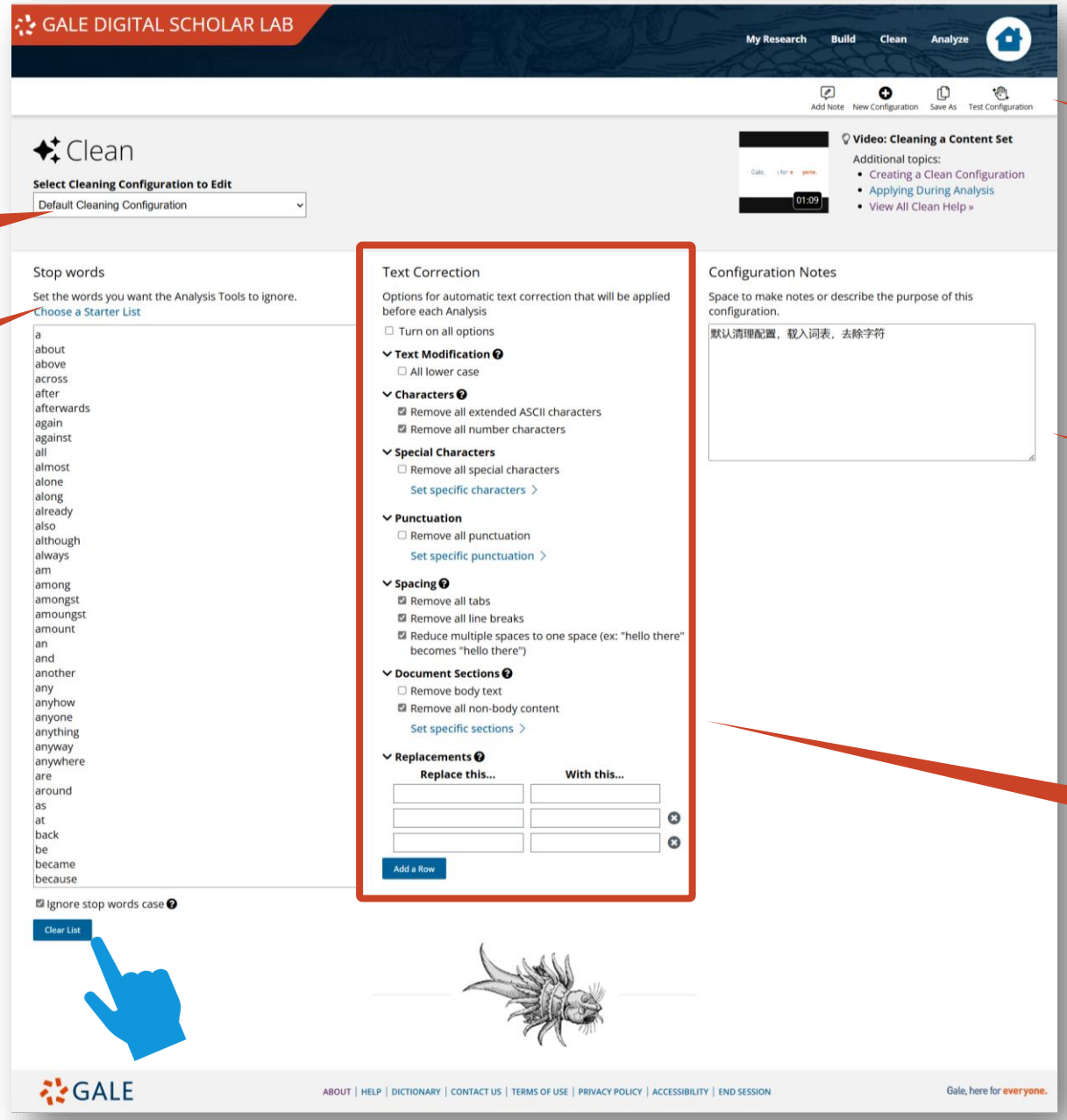
内容集文件夹/内容集

内容集或文件夹详情

Content Sets 内容集



Clean 清理



选择清理配置

停用词表

清理配置管理

清理配置笔记

文本纠正

Clean 清理

更正

文档字符

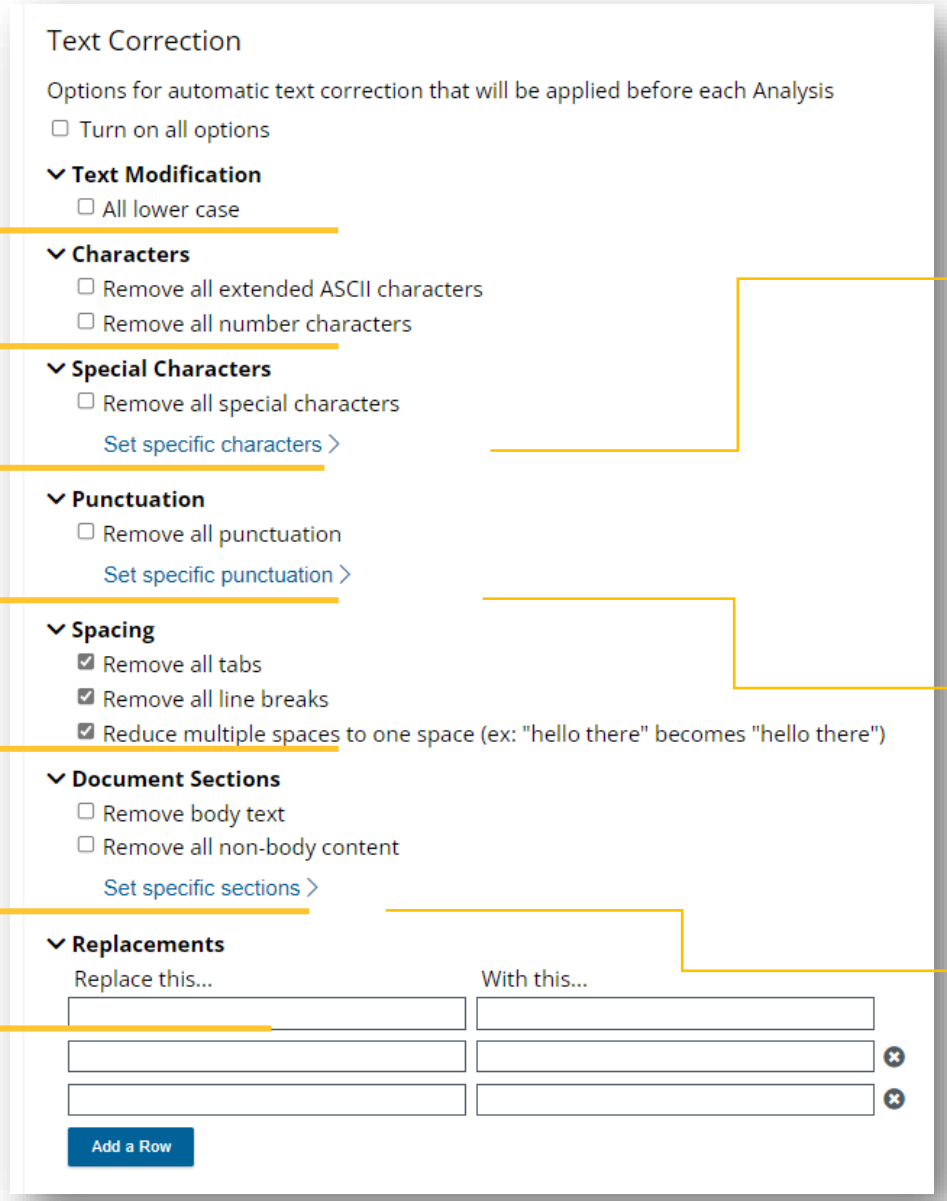
特殊字符

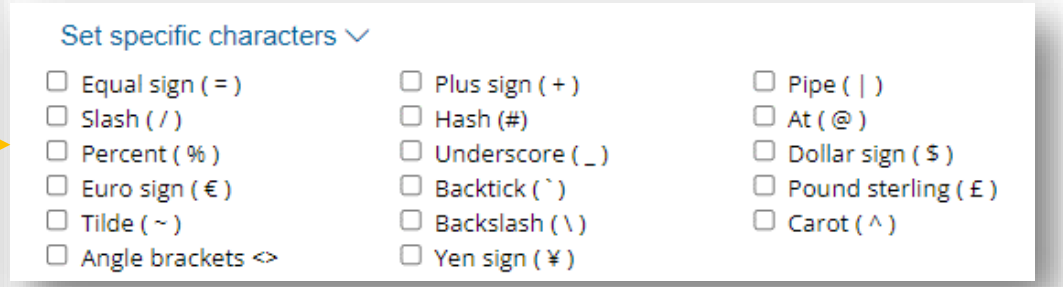
标点

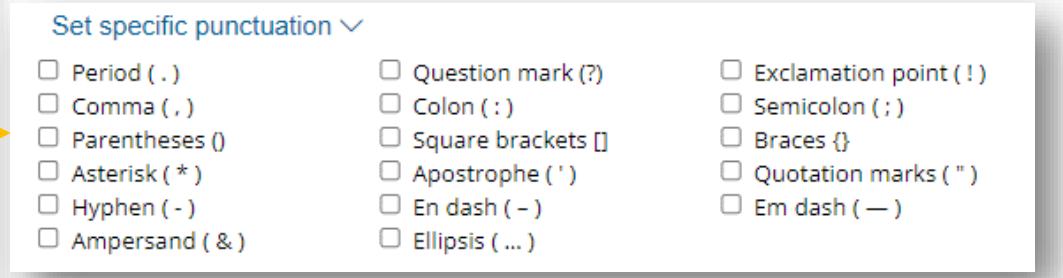
空格

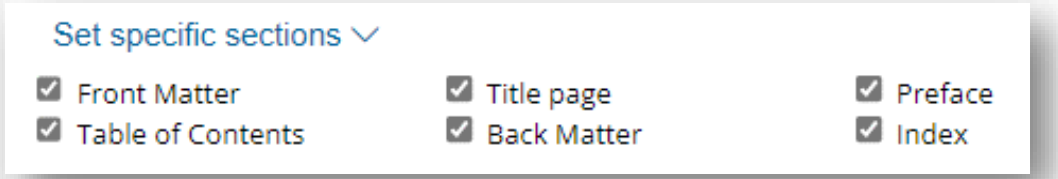
文档拆分

替换

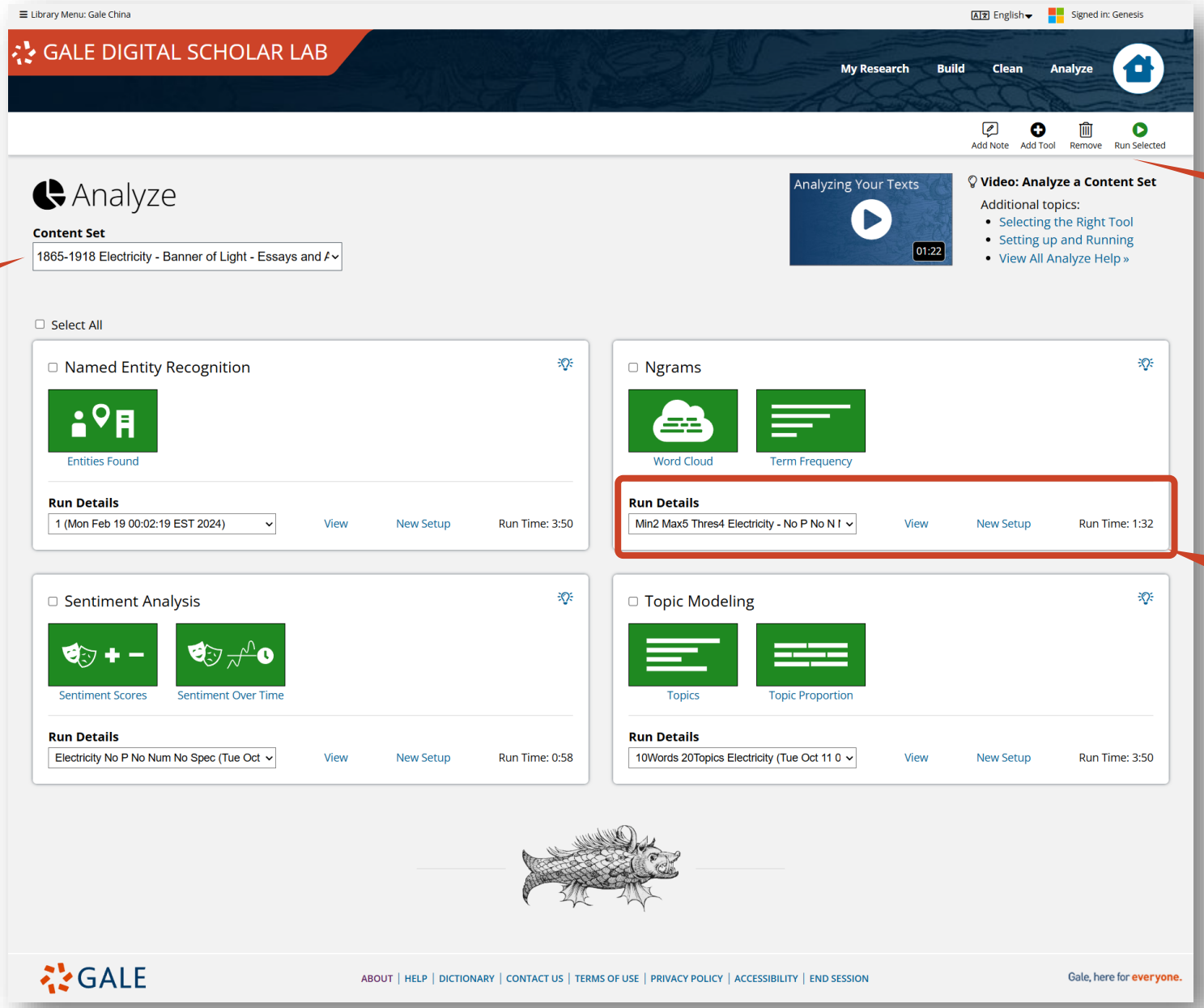








Analyze 分析

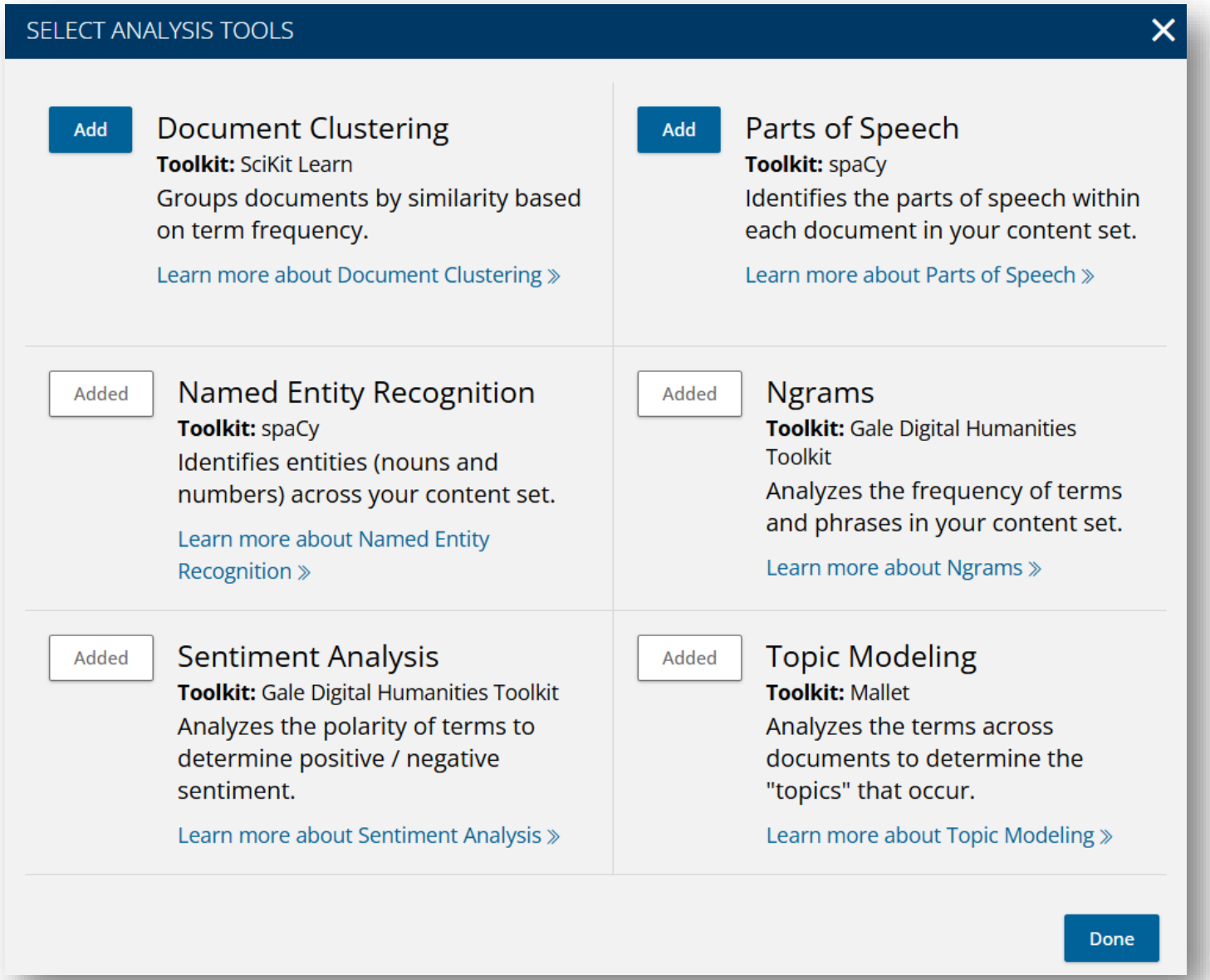


切换内容集

管理分析工具

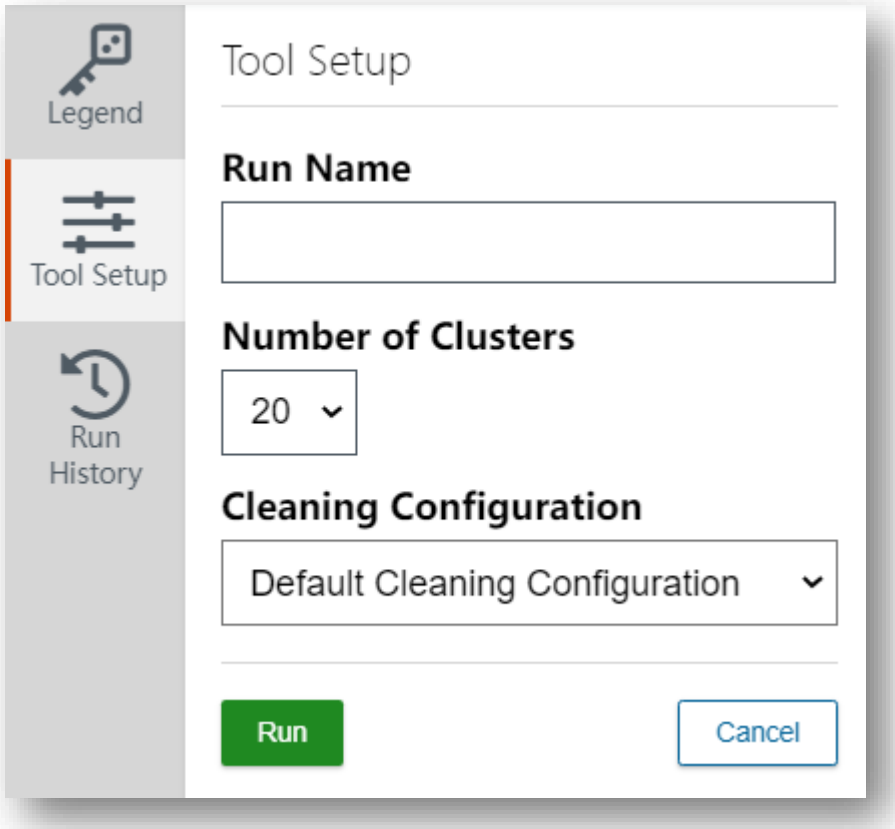
切换分析结果/查看与配置

Analysis Tools 分析工具



- 聚类（Clustering）
- 词性标注（Parts of Speech）
- N元语法（Ngrams）
- 命名实体识别（Named Entity Recognition）
- 情感分析（Sentiment Analysis）
- 主题建模（Topic Modeling）

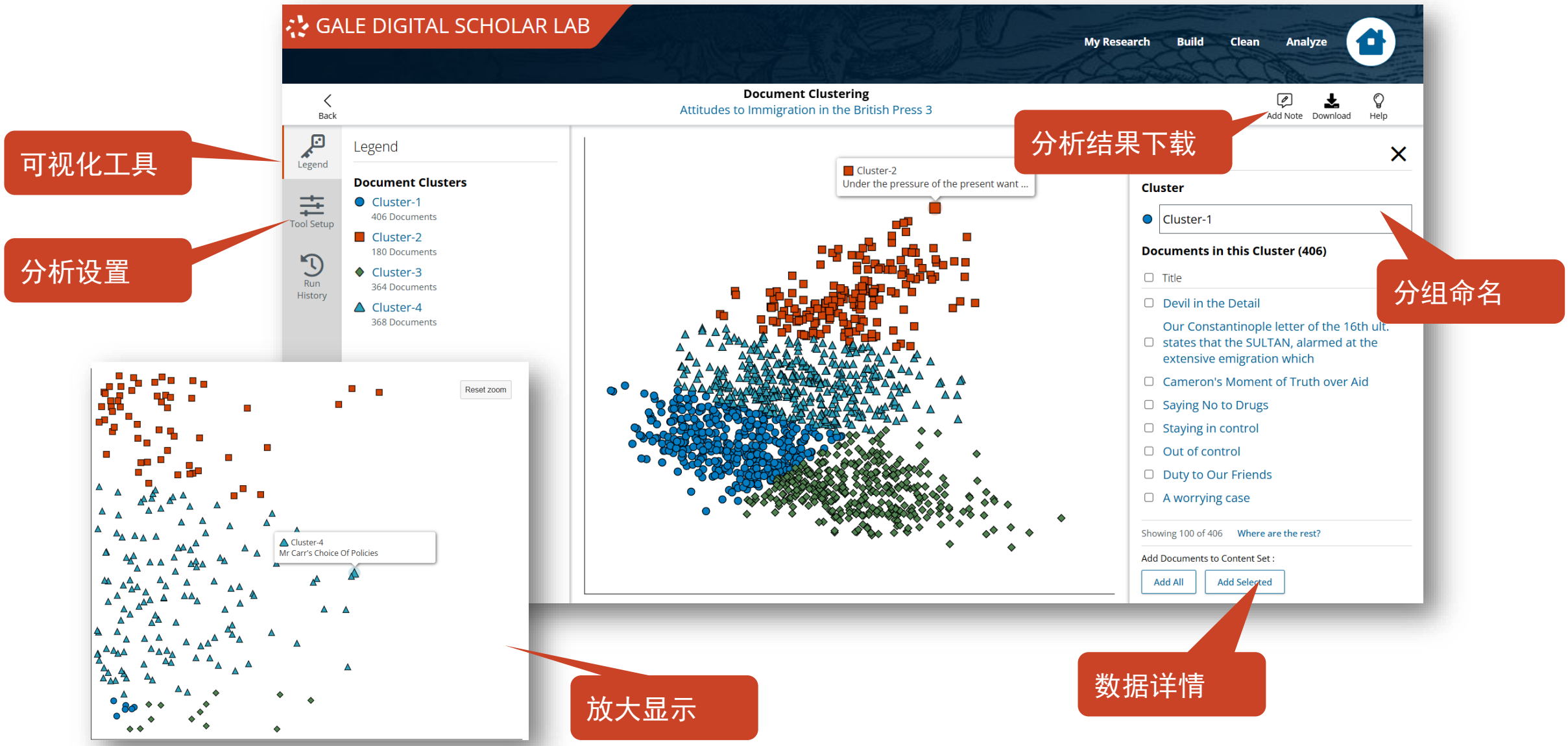
Document Clustering文档聚类



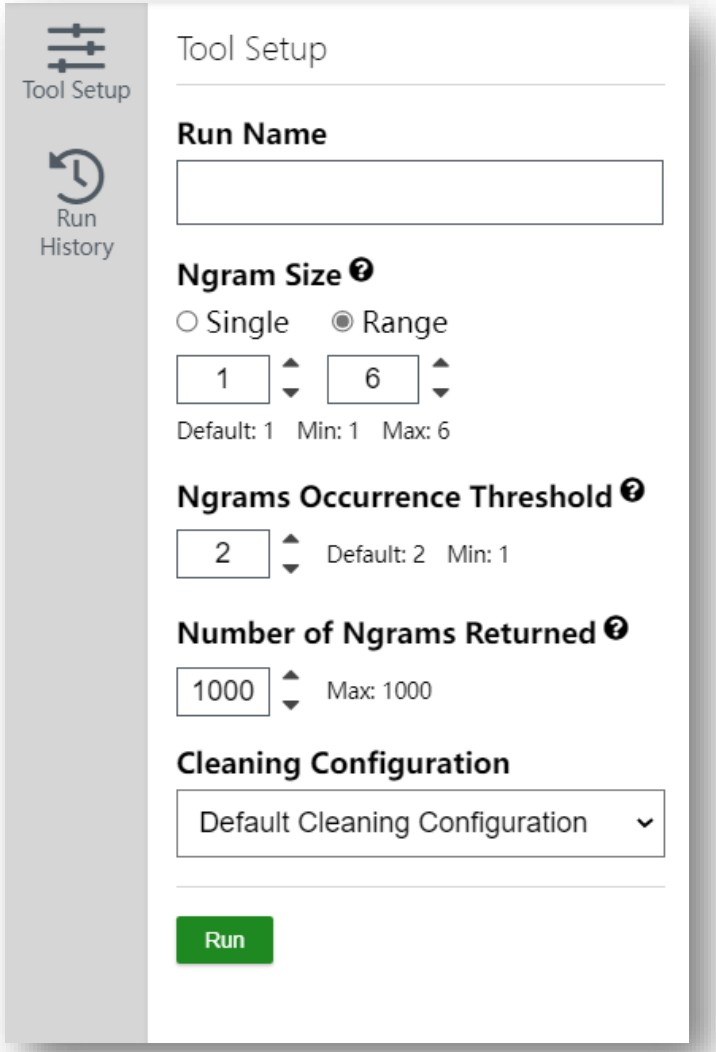
文档聚类

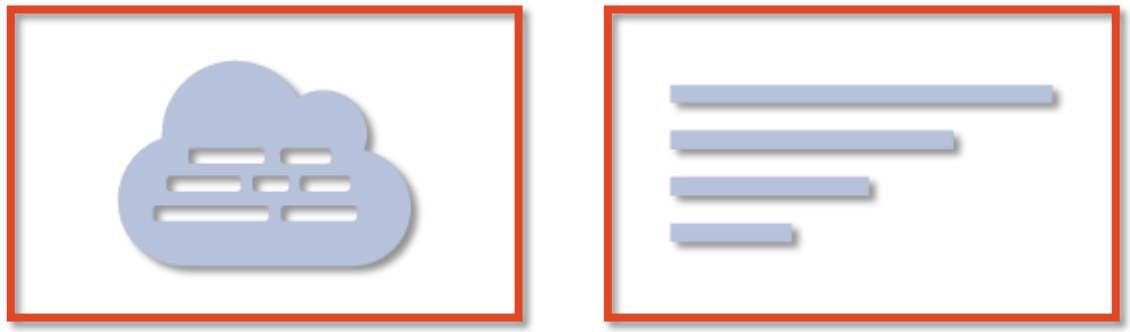
使用统计指标和方法分析内容集中的文档，围绕特定特征或属性对它们进行分组。该算法将所有点聚集到所需数量的簇中，并为属于每个簇的点着色。每个簇代表文档组，这些文档组彼此之间的相似性比与内容集中其他文档的相似性更高。

可视化分析结果


GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Document Clustering
Attitudes to Immigration in the British Press 3
Add Note
Download
Help
可视化工具
分析结果下载
Legend
Tool Setup
Run History
Legend
Document Clusters
Cluster-1
406 Documents
Cluster-2
180 Documents
Cluster-3
364 Documents
Cluster-4
368 Documents
分析设置
Cluster-2
Under the pressure of the present want ...
Cluster
Cluster-1
Documents in this Cluster (406)
Title
Devil in the Detail
Our Constantinople letter of the 16th ult. states that the SULTAN, alarmed at the extensive emigration which
Cameron's Moment of Truth over Aid
Saying No to Drugs
Staying in control
Out of control
Duty to Our Friends
A worrying case
分组命名
Showing 100 of 406
Where are the rest?
Add Documents to Content Set :
Add All
Add Selected
数据详情
Reset zoom
Cluster-4
Mr Carr's Choice Of Policies
放大显示

Ngrams N-元语法



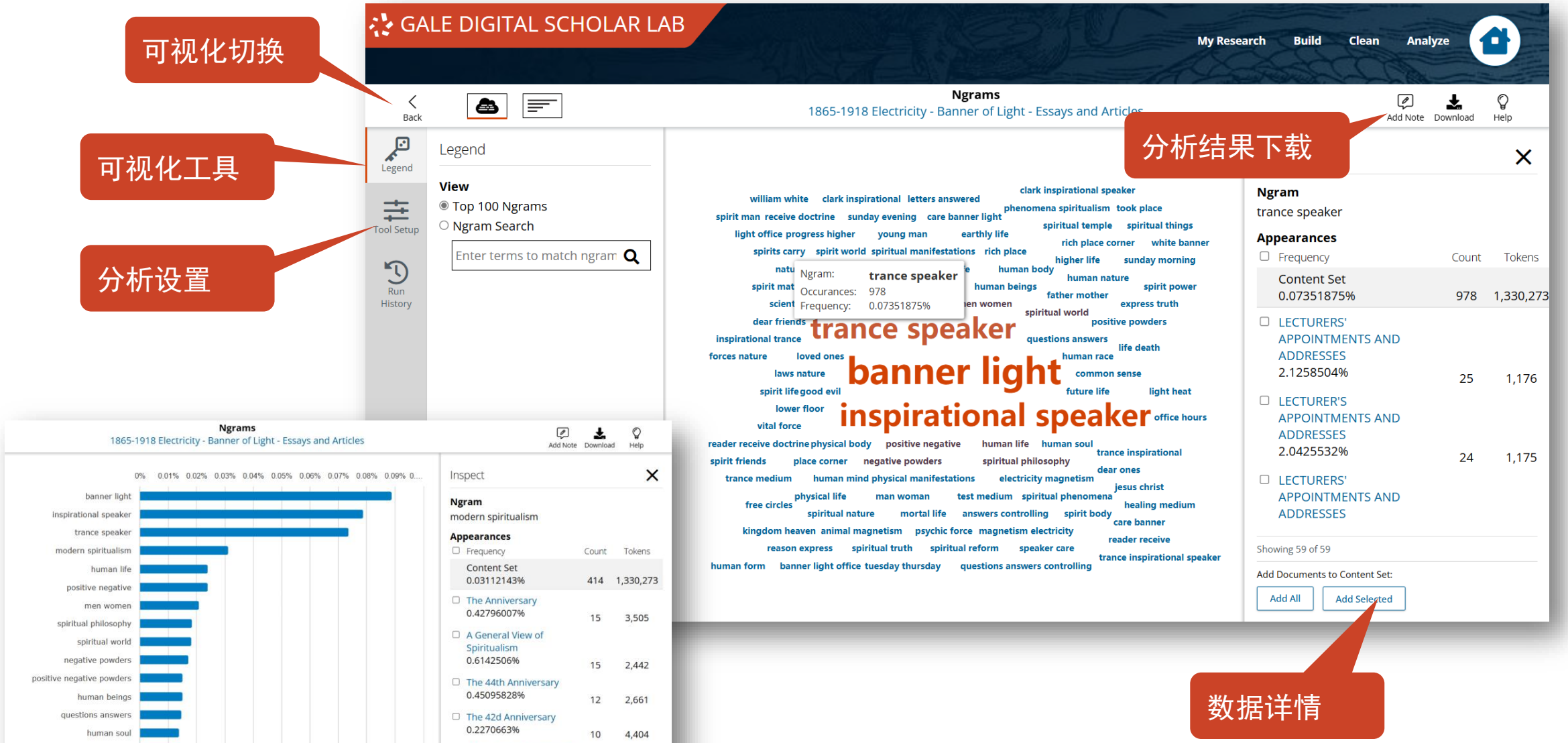


N-元语法

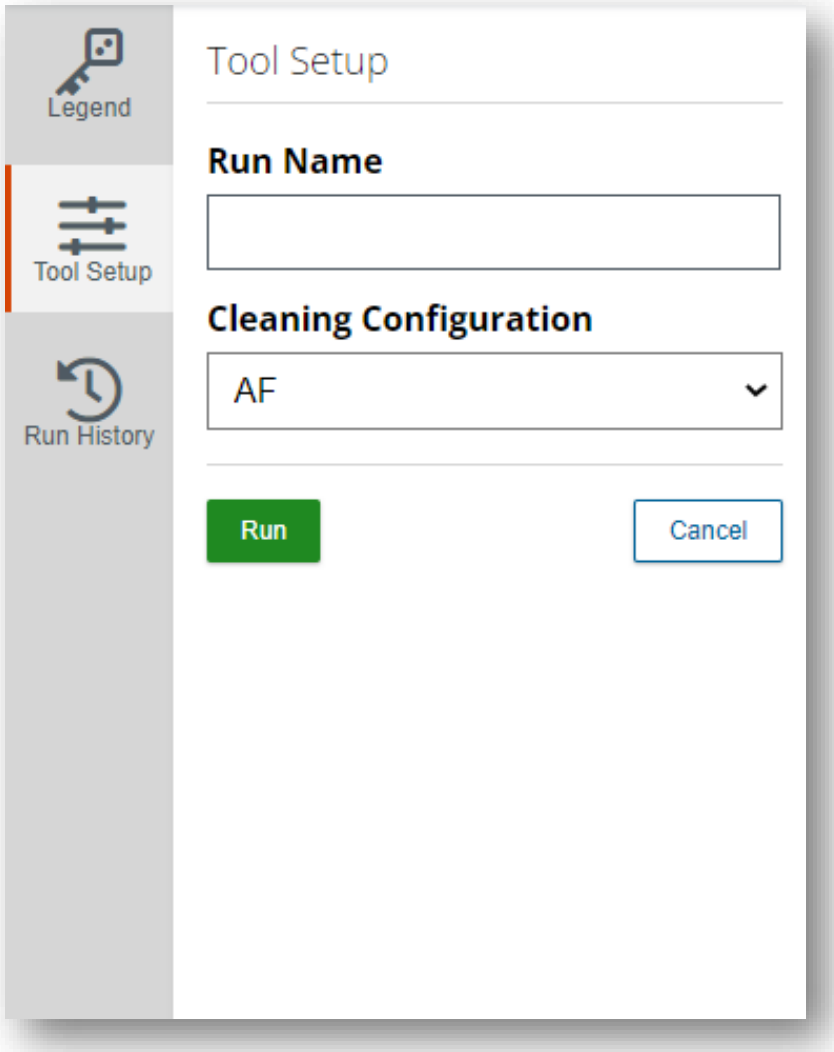
提供了可视化内容集文档中某些单词或标记的频率的方法。

N 表示单词数，单元词有 1 个词，二元词有 2 个词，三元词有 3 个词，等等。

可视化分析结果


可视化切换
可视化工具
分析设置
分析结果下载
数据详情
GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Ngrams
1865-1918 Electricity - Banner of Light - Essays and Articles
Add Note
Download
Help
Legend
Tool Setup
Run History
View
Top 100 Ngrams
Ngram Search
Enter terms to match ngram
Ngram: trance speaker
Occurances: 978
Frequency: 0.07351875%
Ngram
trance speaker
Appearances
Frequency
Count
Tokens
Content Set
0.07351875%
978
1,330,273
LECTURERS' APPOINTMENTS AND ADDRESSES
2.1258504%
25
1,176
LECTURER'S APPOINTMENTS AND ADDRESSES
2.0425532%
24
1,175
LECTURERS' APPOINTMENTS AND ADDRESSES
Showing 59 of 59
Add Documents to Content Set:
Add All
Add Selected
Inspect
Ngram
modern spiritualism
Appearances
Content Set
0.03112143%
414
1,330,273
The Anniversary
0.42796007%
15
3,505
A General View of Spiritualism
0.6142506%
15
2,442
The 44th Anniversary
0.45095828%
12
2,661
The 42d Anniversary
0.2270663%
10
4,404
Showing 100 of 192
Where are the rest?

Parts of Speech词性标注



词性标注

它为用户提供了查看内容集中每个文档中短语构造方式的构建块。此工具可有效创建内容集的词典索引或词表。

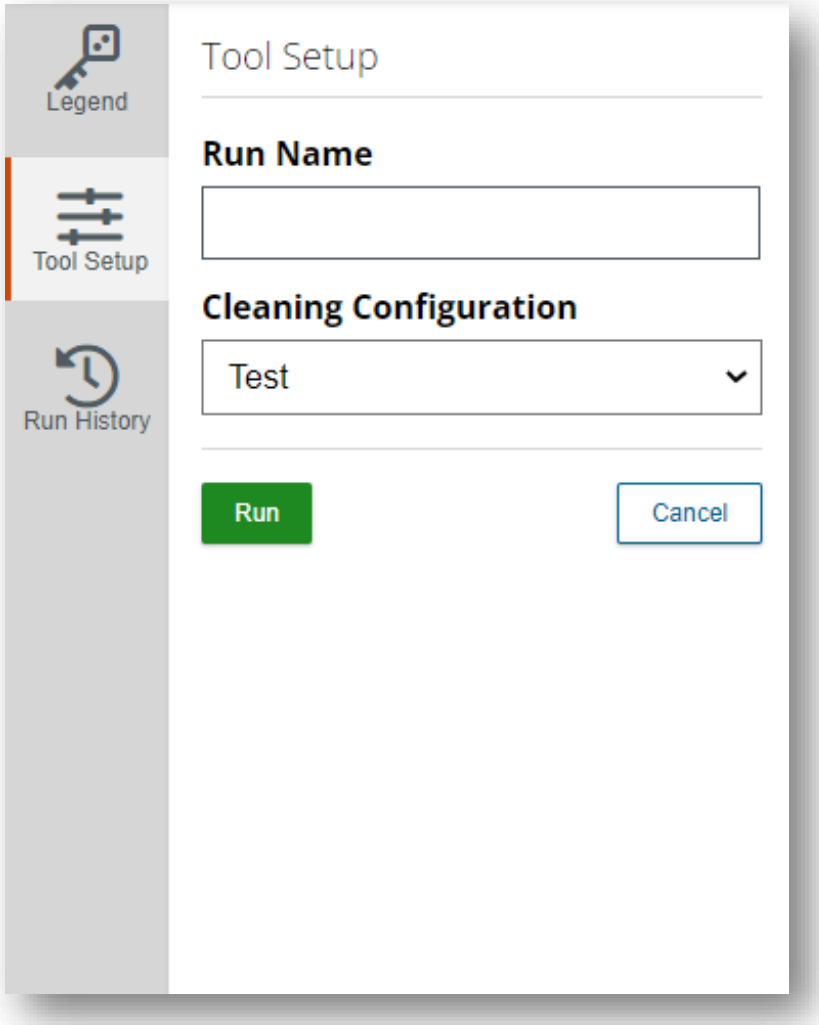
词类包括：连词、代词、专有名词、符号、基数、形容词、连词、空格、限定词、副词、标点符号、助词、动词、感叹词、名词等。

可视化分析结果


GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Parts of Speech
Attitudes to Immigration in the British Press 3
Add Note
Download
Help
Legend
Tool Setup
Run History
Legend
Parts of Speech
Click to toggle categories on and off
NOUN
INTJ
PART
PROPN
CCONJ
PUNCT
PRON
SCONJ
NUM
VERB
DET
SYM
ADV
ADP
SPACE
ADJ
AUX
OTHER
Group Data By
Author
4 of 4 selected
Change selection
Graphs Shown: 10
Grid Columns: 2
David Wood
2 documents | Tagged 977 Parts of Speech
Our Paris Correspondent
1 documents | Tagged 249 Parts of Speech
Unknown Author
1,314 documents | Tagged 317,820 Parts of Speech
Percent (%)
可视化工具
分析设置
分析结果下载


切换显示
Legend
Parts of Speech
Click to toggle categories on and off
Group Data By
Publication
Content Set
Author
Publisher
Publication
Content Type
Document Type
Source Library
Daily Mail
Daily Mail Atlantic Edition
The Times

Named Entity Recognition命名实体识别

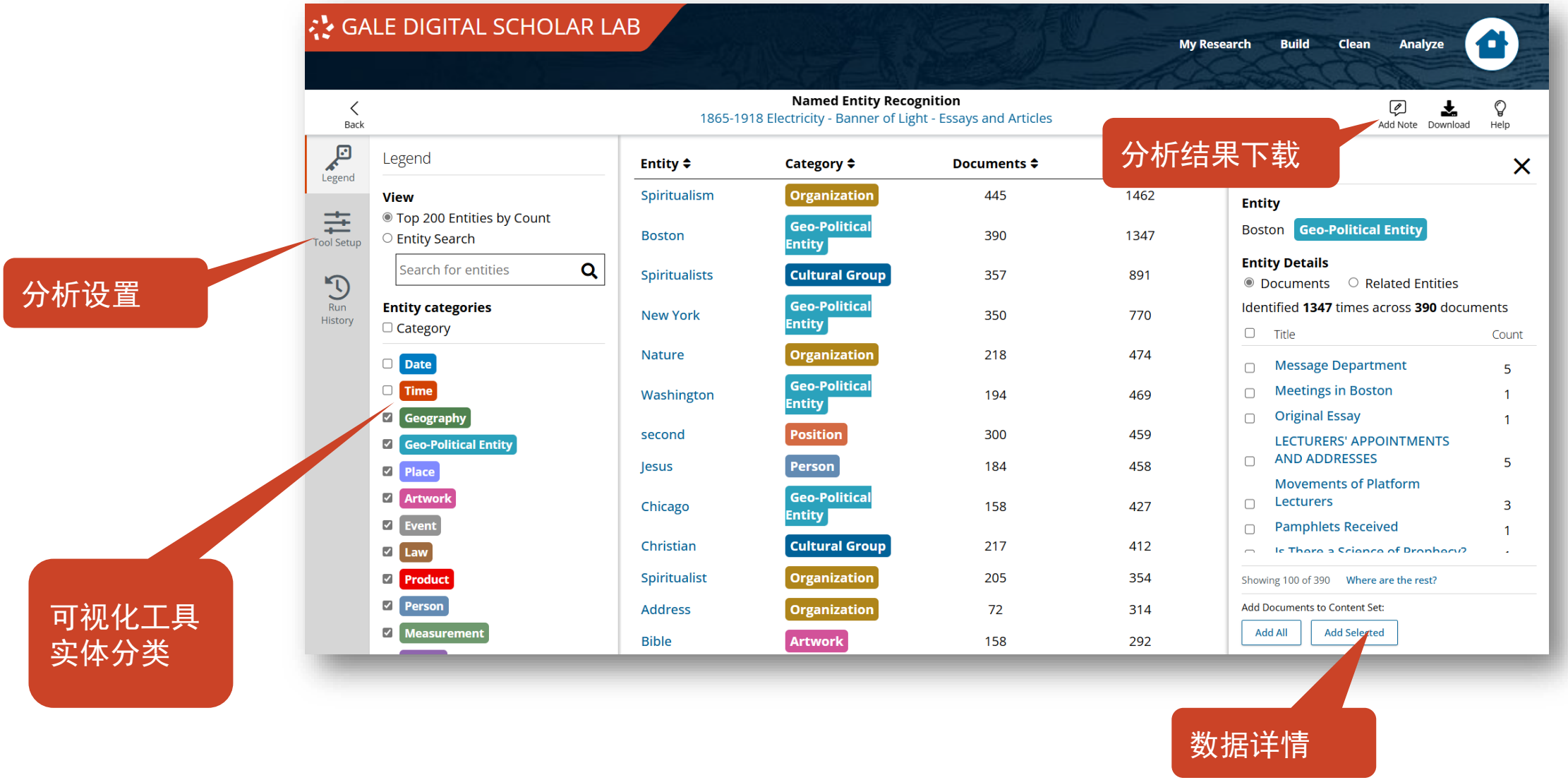


命名实体识别

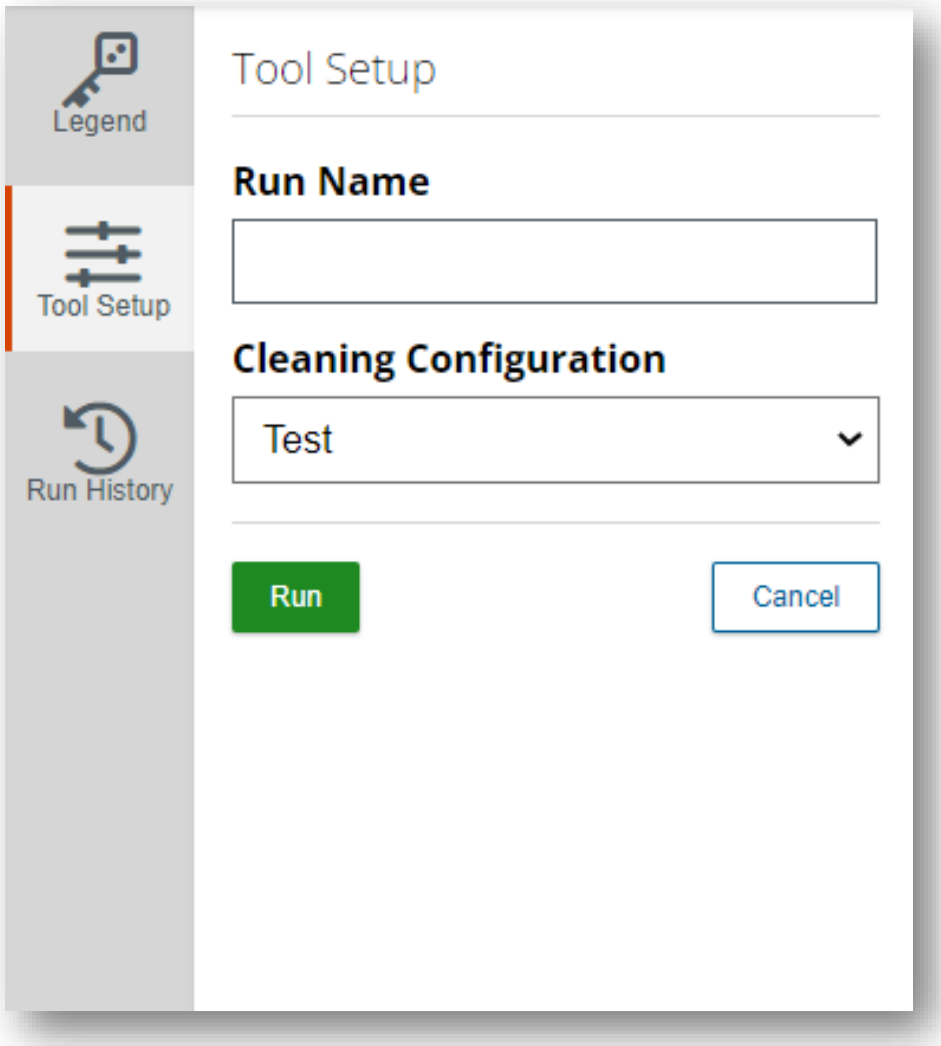
命名实体识别（NER）是一种自然语言处理方法，旨在将内容集中的每个术语识别并分类为特定的实体类别或“类”。

包括：数字、日期、事件、地方、地缘政治实体、语言、法律、地理、钱、文化团体、位置、组织、百分比、人、产品、测量、时间、艺术品等

可视化分析结果


GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Named Entity Recognition
1865-1918 Electricity - Banner of Light - Essays and Articles
Add Note
Download
Help
Legend
Tool Setup
Run History
Legend
View
Top 200 Entities by Count
Entity Search
Search for entities
Entity categories
Category
Date
Time
Geography
Geo-Political Entity
Place
Artwork
Event
Law
Product
Person
Measurement
Entity
Category
Documents
Spiritualism Organization 445 1462
Boston Geo-Political Entity 390 1347
Spiritualists Cultural Group 357 891
New York Geo-Political Entity 350 770
Nature Organization 218 474
Washington Geo-Political Entity 194 469
second Position 300 459
Jesus Person 184 458
Chicago Geo-Political Entity 158 427
Christian Cultural Group 217 412
Spiritualist Organization 205 354
Address Organization 72 314
Bible Artwork 158 292
Entity
Boston Geo-Political Entity
Entity Details
Documents
Related Entities
Identified 1347 times across 390 documents
Title
Count
Message Department 5
Meetings in Boston 1
Original Essay 1
LECTURERS' APPOINTMENTS AND ADDRESSES 5
Movements of Platform Lecturers 3
Pamphlets Received 1
Showing 100 of 390
Where are the rest?
Add Documents to Content Set:
Add All
Add Selected
分析结果下载
分析设置
可视化工具
实体分类
数据详情

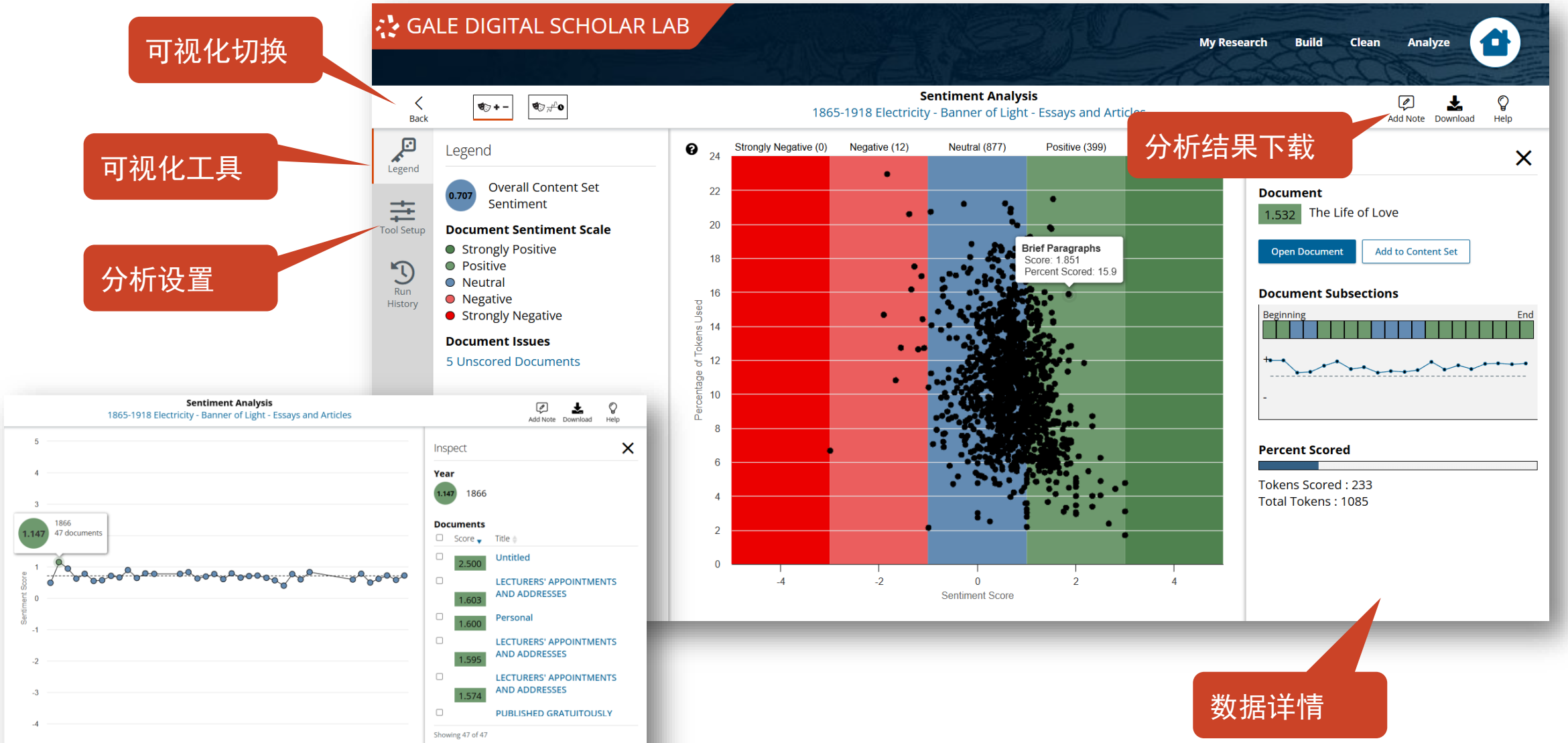
Sentiment Analysis情感分析



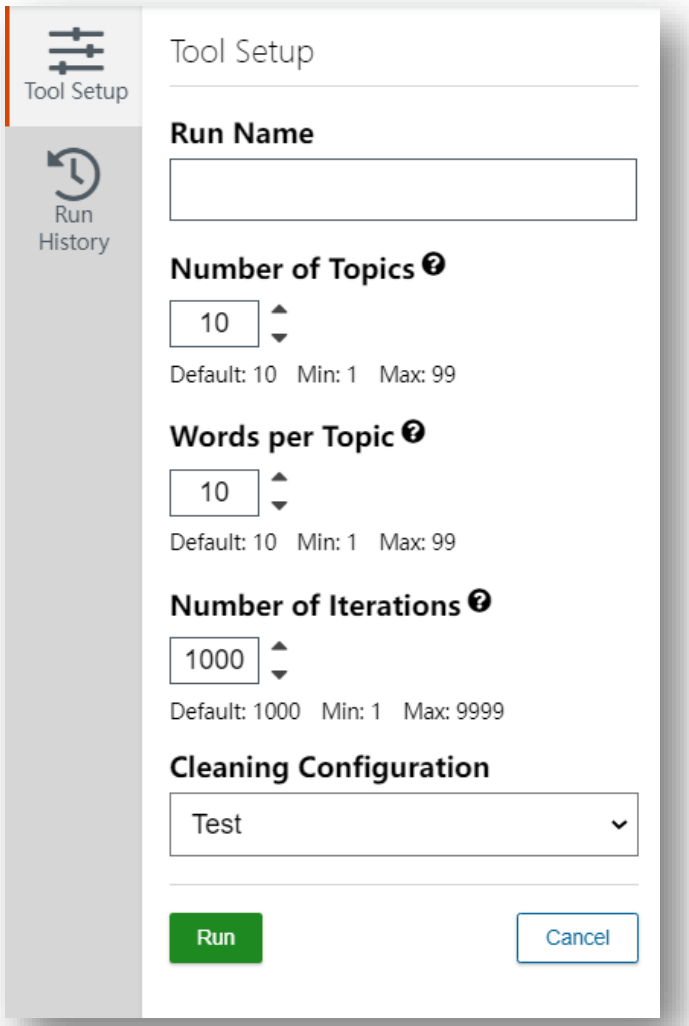
情感分析

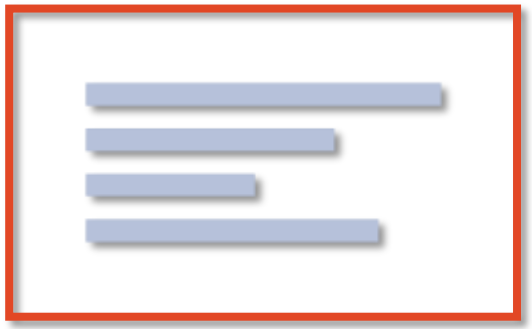
情感分析使用户能够直观地了解其内容集中的积极和消极情感意义。情感意义分析为用户创造了机会，让他们探索作者如何以及为何以积极或消极的方式与主题或事物相关。

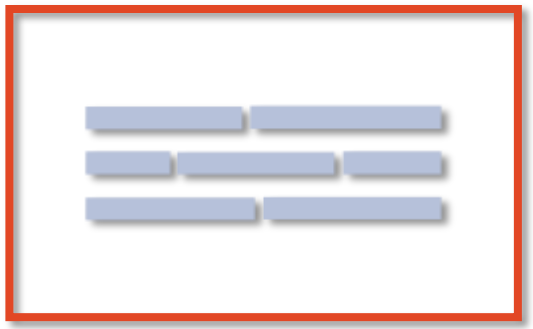
可视化分析结果


可视化切换
可视化工具
分析设置
分析结果下载
数据详情
GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Sentiment Analysis
1865-1918 Electricity - Banner of Light - Essays and Articles
Add Note
Download
Help
Legend
Tool Setup
Run History
Legend
0.707
Overall Content Set Sentiment
Document Sentiment Scale
Strongly Positive
Positive
Neutral
Negative
Strongly Negative
Document Issues
5 Unscored Documents
Strongly Negative (0)
Negative (12)
Neutral (877)
Positive (399)
Brief Paragraphs
Score: 1.851
Percent Scored: 15.9
Percentage of Tokens Used
Sentiment Score
Document
1.532
The Life of Love
Open Document
Add to Content Set
Document Subsections
Beginning
End
Percent Scored
Tokens Scored : 233
Total Tokens : 1085
Inspect
Year
1.147
1866
Documents
Score
Title
2.500
Untitled
1.603
LECTURERS' APPOINTMENTS AND ADDRESSES
1.600
Personal
1.595
LECTURERS' APPOINTMENTS AND ADDRESSES
1.574
LECTURERS' APPOINTMENTS AND ADDRESSES
PUBLISHED GRATUITOUSLY
Showing 47 of 47
Add Documents to Content Set:
Add All
Add Selected
1866
47 documents
Sentiment Score

Topic Modeling主题建模



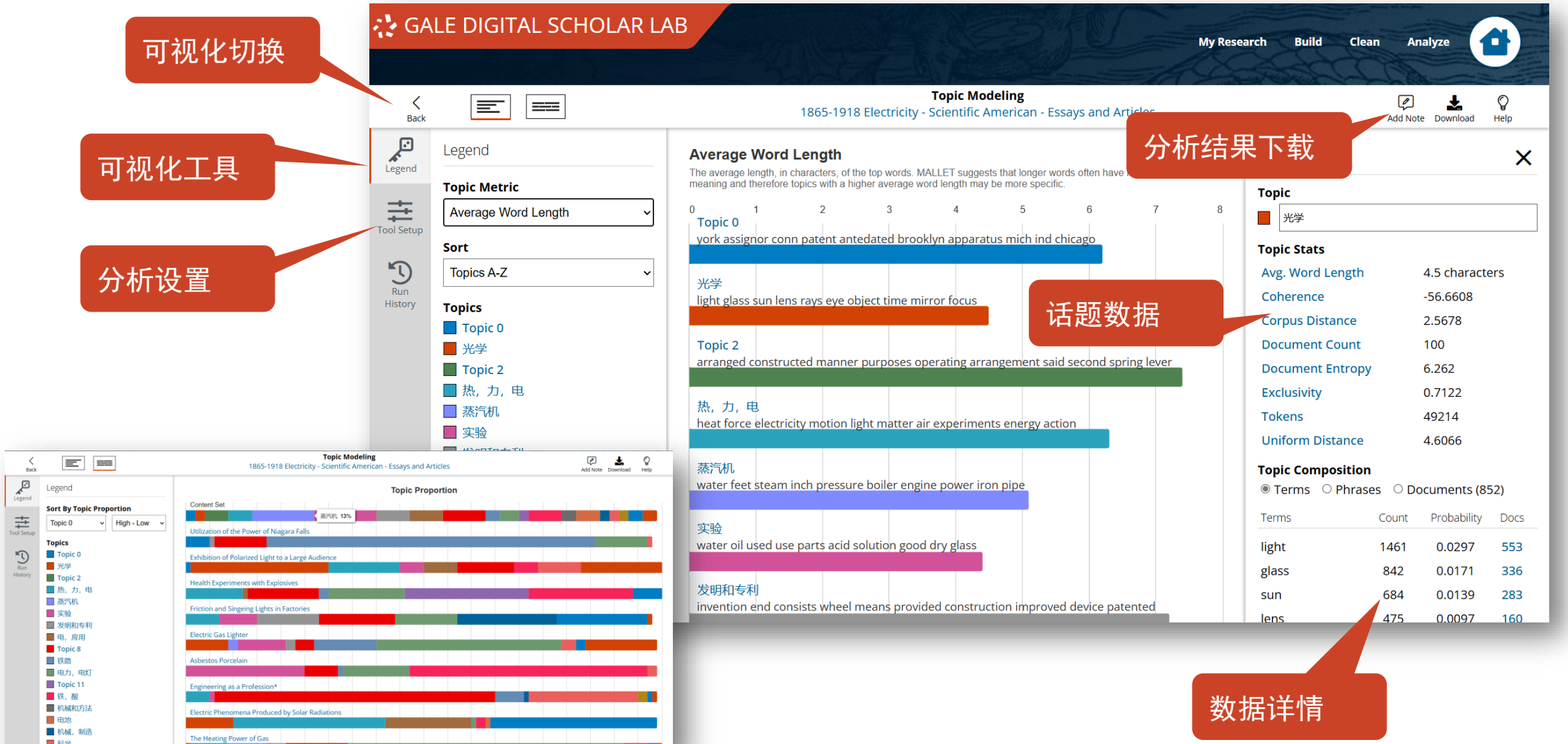




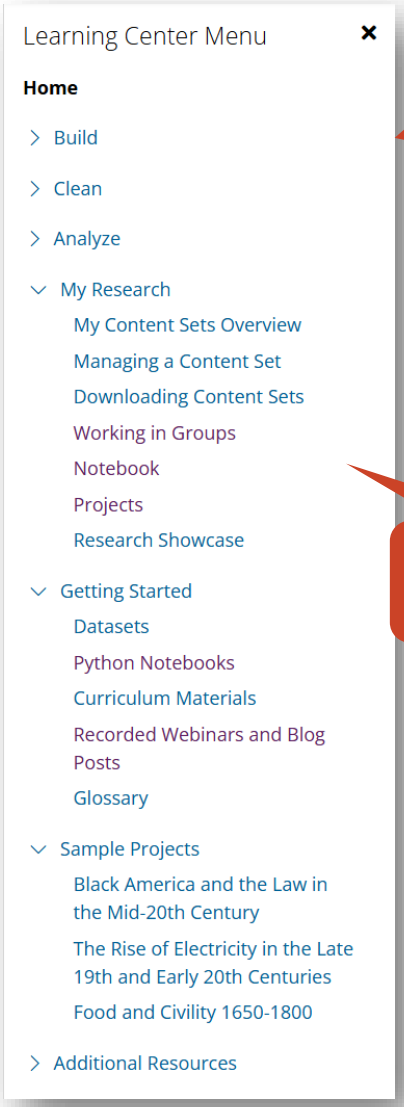
主题建模

主题建模是一种文本挖掘方法，它允许用户通过分析 MALLET 分配的高级主题来辨别其内容集中的情况，既它允许用户在内容集中的文档中发现统计上更有可能彼此靠近出现的词组。

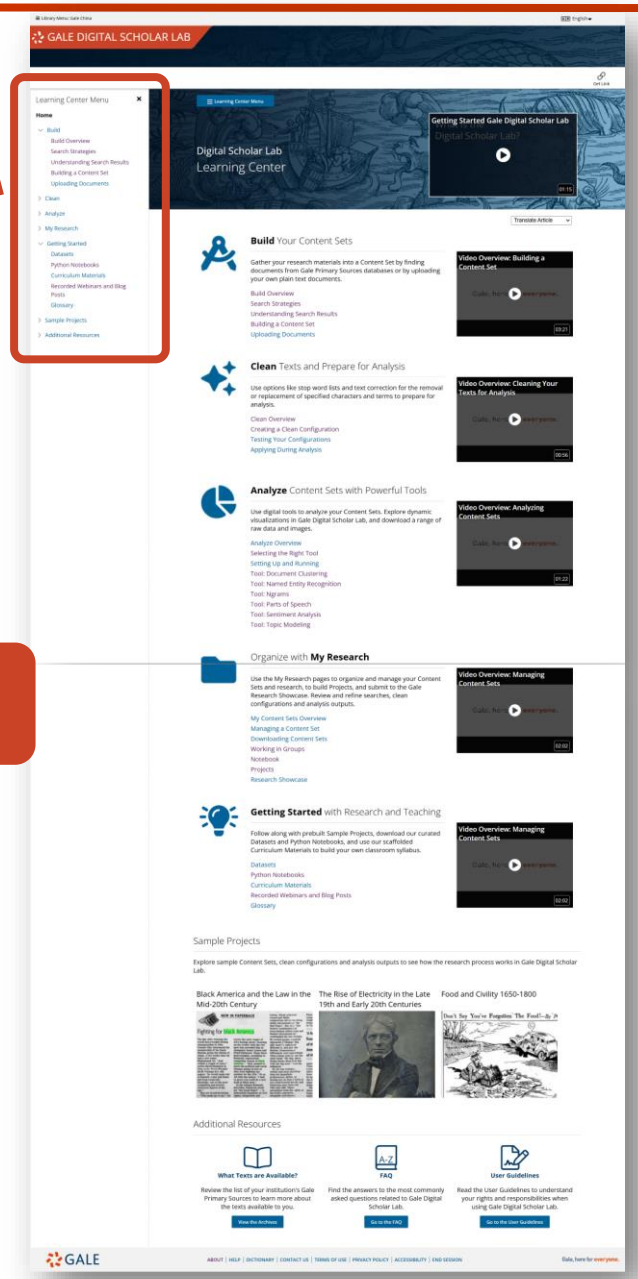
可视化分析结果


可视化切换
可视化工具
分析设置
分析结果下载
话题数据
数据详情
GALE DIGITAL SCHOLAR LAB
My Research
Build
Clean
Analyze
Back
Topic Modeling
1865-1918 Electricity - Scientific American - Essays and Articles
Add Note
Download
Help
Legend
Tool Setup
Run History
Topic Metric
Average Word Length
Sort
Topics A-Z
Topics
Topic 0
光学
Topic 2
热，力，电
蒸汽机
实验
Average Word Length
The average length, in characters, of the top words. MALLET suggests that longer words often have
meaning and therefore topics with a higher average word length may be more specific.
0 1 2 3 4 5 6 7 8
Topic 0
york assignor conn patent antedated brooklyn apparatus mich ind chicago
光学
light glass sun lens rays eye object time mirror focus
Topic 2
arranged constructed manner purposes operating arrangement said second spring lever
热，力，电
heat force electricity motion light matter air experiments energy action
蒸汽机
water feet steam inch pressure boiler engine power iron pipe
实验
water oil used use parts acid solution good dry glass
发明和专利
invention end consists wheel means provided construction improved device patented
Topic
光学
Topic Stats
Avg. Word Length 4.5 characters
Coherence -56.6608
Corpus Distance 2.5678
Document Count 100
Document Entropy 6.262
Exclusivity 0.7122
Tokens 49214
Uniform Distance 4.6066
Topic Composition
Terms Phrases Documents (852)
Terms Count Probability Docs
light 1461 0.0297 553
glass 842 0.0171 336
sun 684 0.0139 283
lens 475 0.0097 160
Topic Proportion
Sort By Topic Proportion
Content Set
Utilization of the Power of Niagara Falls
Exhibition of Polarized Light to a Large Audience
Health Experiments with Explosives
Friction and Singeing Lights in Factories
Electric Gas Lighter
Asbestos Porcelain
Engineering as a Profession*
Electric Phenomena Produced by Solar Radiations
The Heating Power of Gas

Learning Center 学习中心



学习中心目录



- 建立内容集
- 清理文本并准备分析
- 使用功能强大的工具分析内容集
- 整理我的研究
- 开始研究和教学
- 项目案例

更多信息欢迎访问：

www.gale.com

扫描二维码关注Gale官方微信

GALE