

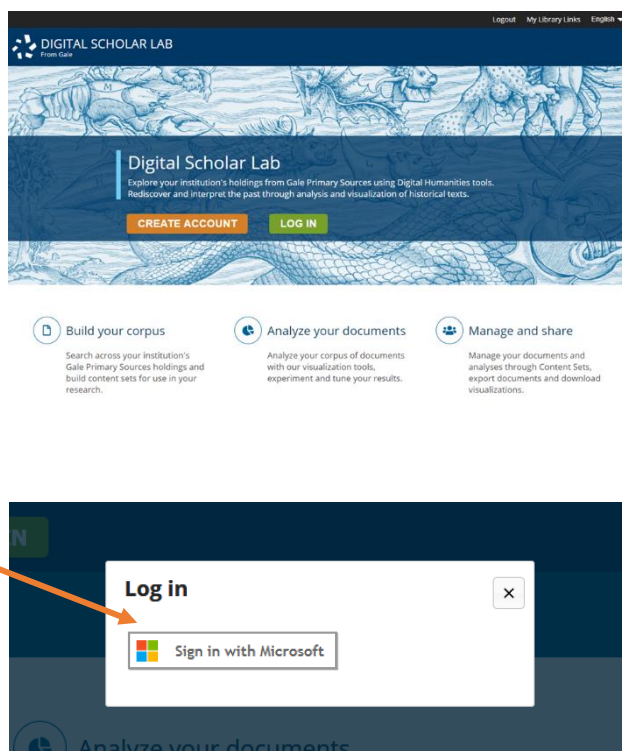


Gale 数字学术实验室——操作指南

准备开始——登录

你的图书馆或院系会提供你登录链接。

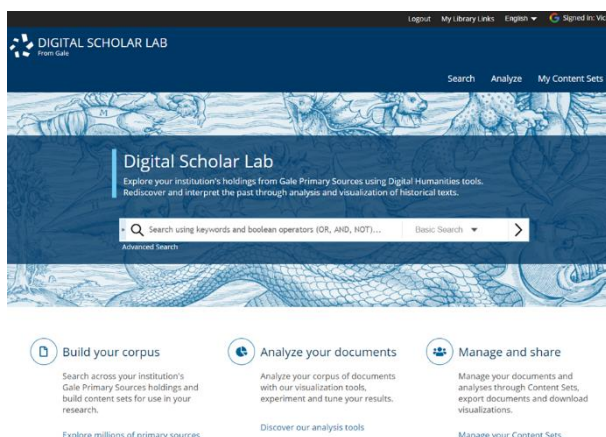
你需要用你个人的微软账户登录，创建你在 Gale 数字学术实验室平台的个人工作账户。



主页——概览

登录后，您将会进入该页面。在此，你可以看到构成实验室平台项目流程的主要组成部分。

整个流程开始于通过数据库形式的检索功能构建你的语料库（个性化档案库或内容集）。随后你可以利用各种文本挖掘分析工具分析你的内容集，最后进入桌面视图，在此管理、组织和分享你在实验室平台中完成的研究。



Gale数字学术实验室——操作指南

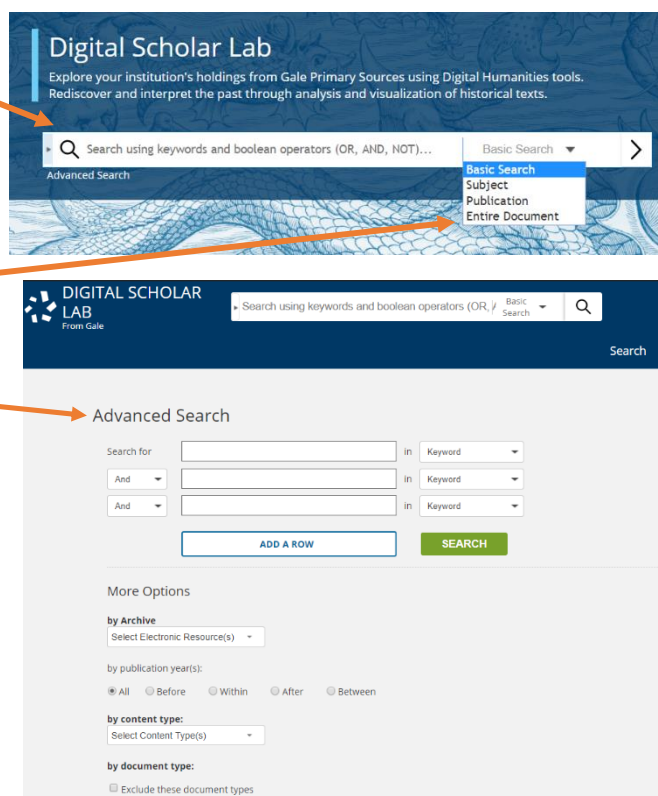
检索

你可以从主页开始检索，此处的**基本检索**选项包括关键词检索及利用“or”、“and”、“not”等逻辑运算符的组合检索。

下拉菜单提供的选项包括按主题或出版物检索，或在整个文档中检索。

你也可以进入**高级检索（Advanced Search）**页面，使用更多检索功能。

高级检索链接位于基本检索框的下方和页面上方菜单中。



检索：建立个性化档案库

建立你的数据集

开始建立你自己的内容集的步骤非常简单。**Gale**的原始档案库内容是实验室平台的基石，能够更快创建个性化档案库正是它的与众不同之处。

输入基本检索的关键词，例如“全球变暖和气候变化”，然后查看你的检索结果。



Gale数字学术实验室——操作指南

检索结果页面

从检索结果页面开始构建你自己的个性化档案库。

你可以立即看到检索返回了多少篇文档，并向下翻页，查看这些内容。

在检索结果中，你可以快速查看每篇文档的详细元数据。其中包括你想要了解的关于每篇文档的多种信息（例如该典藏来源何处、出版日期、作者等）。

你可以看到每篇文档OCR的前几行，以及OCR置信度评级。

OCR置信度是OCR引擎对自己进行文本翻译的确信度。OCR置信度并不等同于OCR准确度。例如，一篇文档可能有很高的准确度，但置信度评分较低，这可能是因为文档中的图片造成的置信度偏差。

利用检索结果页面左侧的边栏，你可以进一步限定检索结果。根据你想要构建的特定内容集选择相关的数据库或主题。

The screenshot shows the Gale Digital Scholar Lab interface. At the top, there's a search bar with the query "global warming and climate change". Below the search bar, the results are displayed in a list format. On the left, there's a sidebar with filters for "Limit Search by:" including Archives, Content Type, and Document Type. The main content area shows two results: "Snaps from space check Earth's health" and "Winter squeeze". Each result includes the author, OCR confidence, a snippet of the text, and a table of metadata (Archive, Source, Library, Content Type, Document, Type, Publication, Date, Publisher). Arrows from the text blocks point to specific elements: one points to the metadata table for "Snaps from space check Earth's health", another points to the OCR confidence and snippet for "Winter squeeze", and a third points to the "Limit Search by:" sidebar.

ACTIVE CONTENT SET
Select Content Set
MANAGE | ANALYZE
Search within results
Limit Search by:
Archives
The Independent Digi... (28)
The Times Digital Ar... (12)
International Herald... (11)
Content Type
Periodical (1)
Newspaper (60)
Manuscript (4)
Document Type
Article (5/1)

All Content (65)
Search Terms: Basic Search ("global warming and climate change")
Sort by: Relevance
Select all on page 1 2 NEXT>

Snaps from space check Earth's health
BY Mark Henderson
OCR CONFIDENCE 73%
NEWS NEWS Snaps from space check Earth's health Snaps from space check Earth's health By Mark Henderson Science Correspondent SCIENTISTS are assembling the most detailed map yet of the Earth's surface, using images from...

Winter squeeze
BY David Marley
OCR CONFIDENCE 67%
Winter squeeze Winter squeeze Sir: Gary Grant's frog spawn observation (letter, 10 January) is another example of a trend towards the end of the traditional winter. Winter is being squeezed out by a longer, warmer autumn...

Snaps from space check Earth's health
ARCHIVE The Times Digital Archive
SOURCE Times Newspapers Limited
LIBRARY
CONTENT TYPE Newspaper
DOCUMENT Article
TYPE
PUBLICATION May 7, 2005
DATE
PUBLICATION The Times.
PUBLISHER

Winter squeeze
ARCHIVE The Independent Digital Archive
SOURCE The Independent
LIBRARY
CONTENT TYPE Newspaper
DOCUMENT Editorial
TYPE
PUBLICATION January 11, 2003
DATE
PUBLICATION The Independent.
PUBLISHER

Winter squeeze
BY David Marley
OCR CONFIDENCE 67%
Winter squeeze Winter squeeze Sir: Gary Grant's frog spawn observation (lett 10 January) is another example of a trend towards the end of the traditional winter. Winter is being squeezed out by a longer, warmer autumn...

Limit Search by:
Archives
The Independent Digi... (28)
The Times Digital Ar... (12)
International Herald... (11)
Content Type
Periodical (1)
Newspaper (60)
Manuscript (4)
Document Type
Article (5/1)





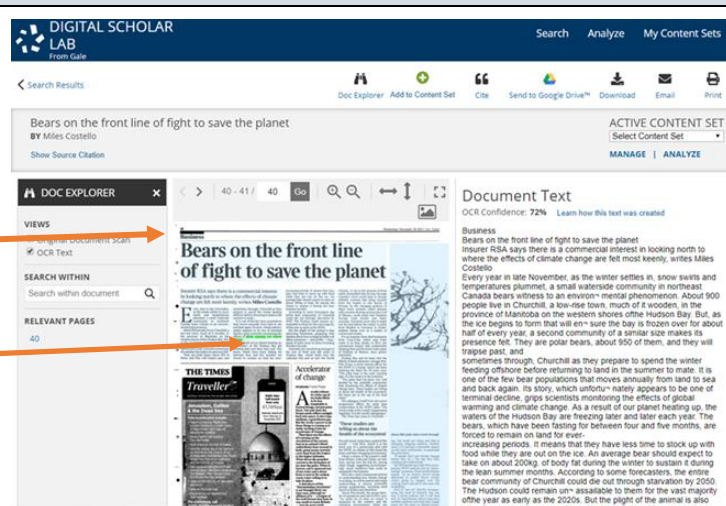
Gale 数字学术实验室——操作指南

文档浏览器

点击文档标题，打开文档浏览器（Doc Explorer）页面。

页面并排显示原始文献资料的图像与它的OCR文本。

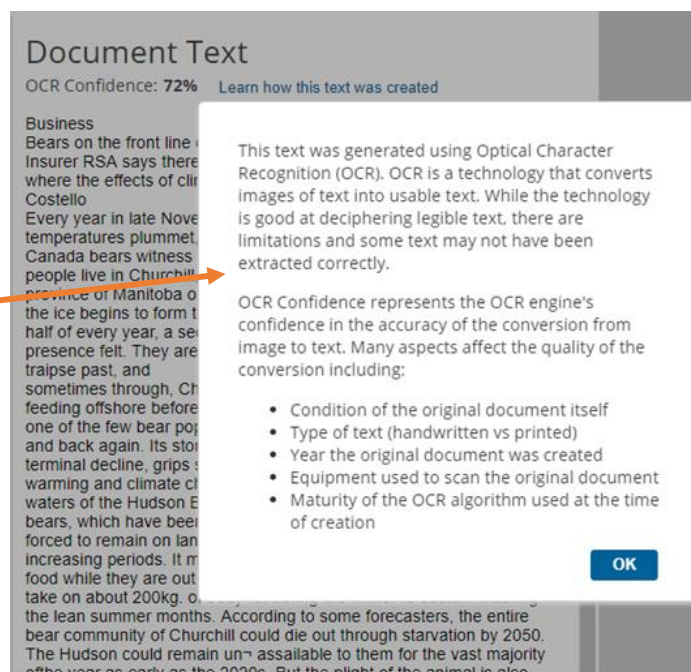
原始文献中会高亮显示用于检索的关键词，以及该文档的来源。



用户可以在这个视图中、文本的上方看到OCR置信水平。

点击“了解OCR文本是如何创建的”（*learn how the OCR text was created*）链接，弹出的窗口中详细解释了这个问题。

此时对OCR文本和原始文献的仔细阅读将对研究者很有帮助，能够快速决定是否需要将这篇文档包括在内容集中，或从内容集中排除。





Gale 数字学术实验室——操作指南

创建内容集

对你选择的检索结果满意后，你可以点击“Add to Content Set”（添加到内容集），弹出你创建的自定义语料库。每篇文档上都会显示你已经将其加入到内容集中。

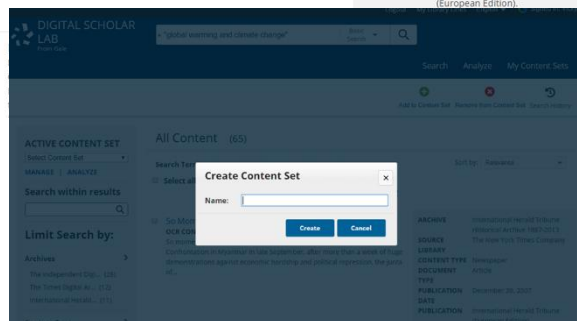
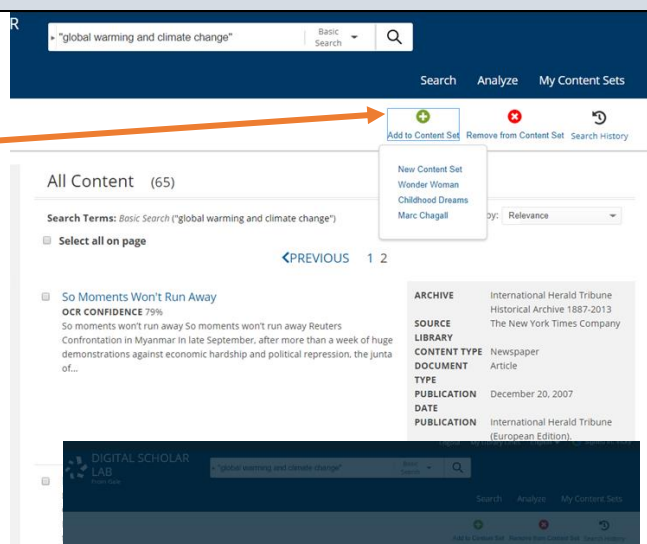
你可以选择添加单个文档。选中多个文档或点击“select all”（选择全部），可添加每个检索页面上的最多100个文档。

你可以选择创建新的内容集，或添加到现有内容集中。这样就建立了你能够管理和保存的文档集合。

每次你登录平台后，都可以看到可供你使用的内容集列表。

你可以在任何时候浏览和编辑你的内容集。

它的一个关键优势在于，你无需重新从头重新构建你的语料库——你可以根据需要不断进行完善和更新。你可以复制内容集，帮助对你的研究项目进行版本管理。



CONTENT SETS

SELECT ALL

Childhood Dreams [Edit](#) [Copy](#)

QUICK STATS Documents: 50 Archives Used: 1

Indicators of Global Warming in Natural World [Edit](#) [Copy](#)

QUICK STATS Documents: 14 Archives Used: 6

Marc Chagall [Edit](#) [Copy](#)

QUICK STATS Documents: 50 Archives Used: 6

Wonder Woman [Edit](#) [Copy](#)

QUICK STATS Documents: 7 Archives Used: 2



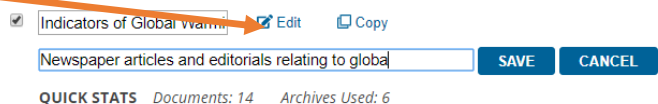
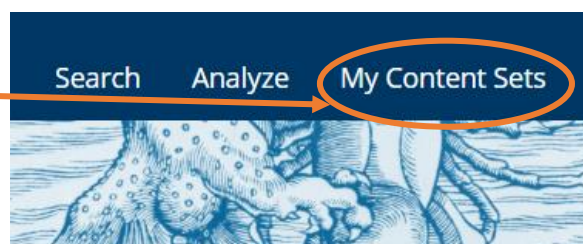


Gale 数字学术实验室——操作指南

编辑内容集

要编辑内容集，你可以从页面上方的“**My Content Sets**”（我的内容集）链接进入，访问你已保存的项目。

点击编辑（**Edit**）按钮，你可以修改内容集名称并添加说明文字。你可以通过检索功能添加或删除文档来管理你的内容集，准备将继续进行分析和可视化处理的资料。



面板视图

点击内容集的名称，就会进入面板视图。

概览（Overview）选项卡提供了对以下信息的总结：

使用的档案库

文档类型

作者

来源图书馆

你也可以选择在此**下载（download）**你的内容集。

Indicators of Global Warming in the Natural World [DOWNLOAD](#)

Newspaper articles and editorials relating to global warming.

Overview

Documents (21)

Analyses

Search History

ARCHIVES USED (6)

The Times Digital Archive (8)
The Independent Digital Archive (8)
Daily Mail Historical Archive, 1896-2004 (1)
The Sunday Times Digital Archive (1)
International Herald Tribune Historical Archive 1887-2013 (1)

[View More](#)

DOCUMENT TYPE (2)

Article (20)
Editorial (1)

AUTHORS (17)

Michael McCarthy (2)
Lewis Smith (2)
Matthew Beard (1)
Mark Nigel Henderson Hawkes (1)
Miles Costello (1)

[View More](#)

SOURCE LIBRARIES (5)

Times Newspapers Limited (10)
The Independent (8)
Associated Newspapers Limited (1)
The New York Times Company (1)
Telegraph Media Group (1)





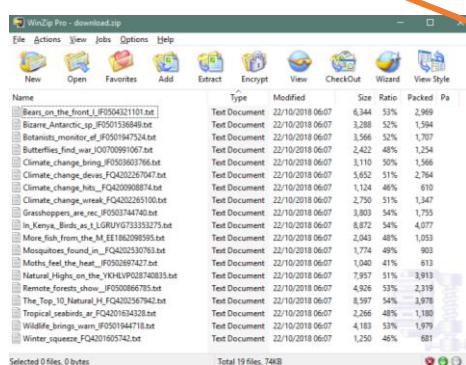
Gale数字学术实验室——操作指南

面板视图（续）

点击下载（download）按钮，你就启动了下载程序。弹出窗口将会打开，详细解释下载参数。

你需要刷新浏览器来更新状态，准备就绪后，下载按钮会变成“download ready”（下载准备就绪）。

点击这个按钮，你会看到和之前类似的信息，说明你可以每次下载最多1,000篇文档，但同时出现了一个绿色的下载按钮。点击后，将会下载包含每篇文档OCR文本的压缩包。



文档（Document）选项卡列出了内容集中的每篇文档，并可点击链接在文档浏览器中查看，并显示出每篇文档的详细元数据。

通过选择筛选选项，可以仅显示部分文档。

Download Content Set

Users may download up to 1,000 documents per session from a selected content set. A sampling of 1,000 documents will be selected for download for content sets that contain more than this threshold. Because of the size, a zip file must first be generated and may take a few minutes.

CONTENT SET

Indicators of Global Warming in the Natural World

Document count: 21

Your download is being generated. Please refresh your browser for a status update.

By clicking null you agree to the [terms of use](#) for this content as agreed to by your licensed institution.

DOWNLOAD READY

Download Content Set

Users may download up to 1,000 documents per session from a selected content set. A sampling of 1,000 documents will be selected for download for content sets that contain more than this threshold. Because of the size, a zip file must first be generated and may take a few minutes.

Download

Last generated: Mon Oct 22 06:04:18 EDT 2018
Document count: 21

CONTENT SET

Indicators of Global Warming in the Natural World

Document count: 21

REGENERATE DOWNLOAD

By clicking Regenerate Download you agree to the [terms of use](#) for this content as agreed to by your licensed institution.

Indicators of Global Warming in the Natural World [DOWNLOAD READY](#)

Newspaper articles and editorials relating to global warming

[Overview](#) [Documents \(21\)](#) [Analyses](#) [Search History](#)

Archive

The Times Digital Archive (3)
The Independent Digital Archive (8)
Daily Mail Historical Archive, 1996-2004 (1)

Document Type

Article (10)

Editorial (1)

Author

Michael McCarthy (2)

James Smith (2)

Matthew Board (1)

Source Library

Times Newspapers Limited (10)

The Independent (8)

Associated Newspapers Limited (1)

Select all on page

Winter squeeze

BY David Marley

OCR CONFIDENCE 67%

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

Winter squeeze

ARCHIVE The Independent Digital Archive
SOURCE The Independent
LIBRARY The Independent
CONTENT Newspaper
TYPE Editorial
DOCUMENT
PUBLICATION January 11, 2003
DATE
PUBLICATION The Independent
PUBLISHER
ARCHIVE The Independent Digital Archive
SOURCE The Independent
LIBRARY The Independent
CONTENT Newspaper
TYPE Article
DOCUMENT
PUBLICATION January 21, 1992
DATE





Gale数字学术实验室——操作指南

分析

分析（**Analyze**）模块是你创建和迭代你的可视化结果的地方。在此，你可以将你的内容集与你想要用来分析你的个性化档案库的分析工具相匹配。

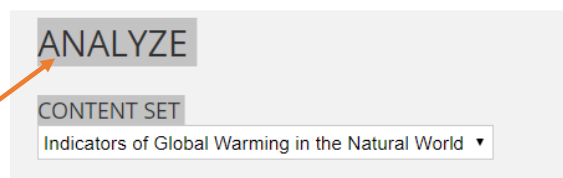
你可以从你构建内容集的检索结果页面进入到分析页面中，或利用导航窗口上的相应链接。

选择好将要分析的内容集后，点击绿色的 **“Add Tool”**（添加工具）按钮，开始选择你想要用来运算你的内容集的分析方法。

你将会看到实验室平台中提供的所有工具。

我们为每一种分析方法都提供了简单易懂的工具说明，揭示了工具的含义及其可能的意义。

了解更多（**Learn More**）链接将打开工具开发者的说明文档，了解更多关于每种特定算法的信息。



Get started by adding an Analysis Tool

ADD TOOL

Analysis Tools

Clustering

Clustering analyzes the documents from a content set using statistical measures and methods to group them around particular features or attributes. This implementation of clustering leverages the k-means algorithm to create clusters of documents according to similar words contained within each document of your content set. [LEARN MORE](#)

ADD

EXAMPLE OUTPUTS



Named Entity Recognition

Named Entity Recognition (NER) recognizes and extracts proper and common nouns from documents using a Parts of Speech tagging method, and outputs them as lists of grouped by entity "type". Some "entity types" available for extraction are: people (including fictional), groups (nationalities, religious, or political), organizations (companies, agencies, institutions, etc.), locations (countries, states, cities), products (objects, vehicles, foods, etc.), works of art (titles of books, songs, etc.), dates (absolute or relative dates or periods), among others. This implementation uses spaCy's Named Entity Recognition model. [LEARN MORE](#)

ADD

EXAMPLE OUTPUTS





Gale数字学术实验室——操作指南

分析（续）

点击添加按钮选择好工具后，返回到分析页面。现在你可以通过选中所有工具并点击“**Run Selected**”（运算选中项）按钮开始分析。

你可以看到每一个工具的运算状态，随着运算的进行状态相应变化。

你也可以**配置（Configure）**每个工具。大部分的分析工具都是可以自定义的，便于用户按自己需要的参数设定工具。所有的工具都有执行分析过程的默认配置，因此无论用户的经验程度如何，都可以生成分析结果。

运算历史（run history）便于你重新生成之前运算中的所有分析结果。

你可以在不同配置参数间切换，查看结果有何不同。你还可以看到每个工具的缩略图，代表每种分析方法产生的不同类型结果。

ANALYZE

CONTENT SET
Indicators of Global Warming in the Natural World

Analysis (6)

Tool	RUN STATUS	CONFIGURATION	RUN HISTORY	ANALYSES
Parts of Speech Tagger	Added	No Configuration	Mon Oct 22 06:57:49 EDT 20	Tabular Data, Series View
Ngram	Added	View Configuration	Mon Oct 22 06:57:47 EDT 20	Tabular Data, Word Cloud
Sentiment Analysis	Added	No Configuration	Mon Oct 22 06:57:50 EDT 20	Tabular Data, Time Series
Named Entity Recognition	Added	No Configuration	Mon Oct 22 06:57:46 EDT 20	Tabular Data

Run In-Progress Tools: 2 In Progress | 7 Completed | Status updates happen every minute

Analysis (6)

Tool	RUN STATUS	CONFIGURATION	RUN HISTORY	ANALYSES
Parts of Speech Tagger	Completed	No Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Series View
Ngram	Completed	View Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Word Cloud
Sentiment Analysis	Completed	No Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Time Series

Run Completed Tools: 11 Completed

Analysis (6)

Tool	RUN STATUS	CONFIGURATION	RUN HISTORY	ANALYSES
Parts of Speech Tagger	Completed	No Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Series View
Ngram	Completed	View Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Word Cloud
Sentiment Analysis	Completed	No Configuration	Mon Oct 22 07:51:20 EDT 20	Tabular Data, Time Series
Named Entity Recognition	Completed	No Configuration	Mon Oct 22 07:51:20 EDT 20	Hierarchical Tree, Tabular Data





Gale数字学术实验室——操作指南

工具

点击每个工具的名称，你可以看到你的分析产生的可视化结果。

点击每种分析方法的工具名称，打开某个工具的浏览器视图。浏览器视图为你呈现了每种分析方法输出哪些可视化结果。

Analysis (6)

☐ Select All

☐ Parts of Speech Tagger

☐ Ngram

☐ Sentiment Analysis

☐ Named Entity Recognition

☐ Topic Modelling

浏览器视图

所有工具都包含一个工具描述链接，其中又包含该分析软件的链接。

例如，如果你点击**Ngram**工具，你将会在左侧边栏看到该工具提供的可视化选项，并可在此访问表格数据。

配置该工具的选项则在主面板中。

在显示的页面中，**Ngram**的大小被调整为最小5，并选择了词云（Word Cloud）作为可视化结果输出。

EXPLORER

Ngram

Tool Description

CONTENT SET

Indicators of Global Warming in the Natural World

RUN HISTORY

Mon Oct 22 08:46:55 EDT 2018

RUN STATUS

completed

TOOL

Configuration

VISUALIZATIONS

Ngram

Word Cloud

Bar Chart

Tabular Data

TOOL CONFIGURATION

Ngram Size

Sets the minimum and maximum size of the ngrams captured in the output (i.e. unigram, bigram, trigram, etc.). The ngram size must be within the range of 1 - 6.

MIN Default: 1

MAX Default: 4

Stop words

A list of words that are excluded from analysis and will not appear in any outputs. All stop words are normalized for case, so different capitalizations do not need to be included (Eg: "above" appears in the list only once, because "Above" would be treated the same as "above").

[Reset To Default](#) [Clear All](#)

a

about

above

warmer weather has led to these studies are telling us about the health of the ecosystem per person including inter national as the effects of global warming

the effects of global warming

person including inter national flights are telling us about the health of the ecosystem the 100 metres up the mountain studies are telling us about the





Gale数字学术实验室——操作指南

浏览器视图及工具（续）

你可以返回分析页面查看利用其它工具生成的可视化结果。

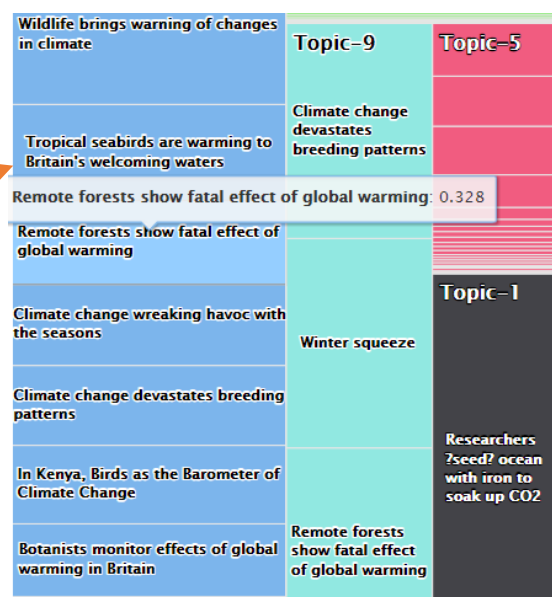
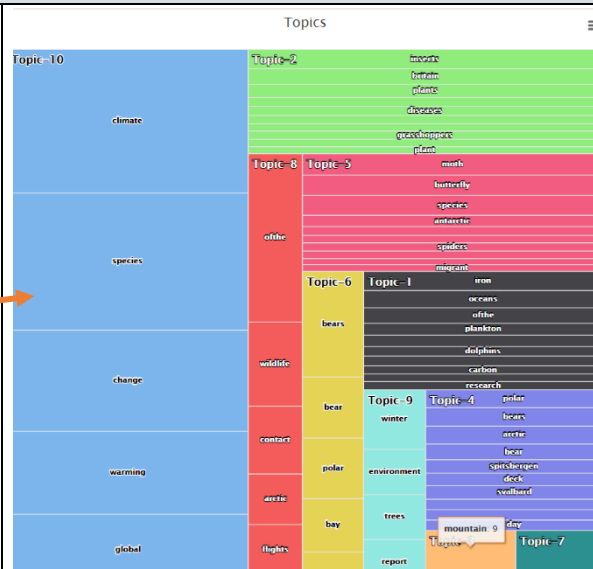
如果你选择了**Topic Modelling**（主题建模）工具，你将会看到多种不同形式的可视化结果。

例如，右图的可视化结果呈现的是按主题（Topic）显示的**树状图（Tree Map）**。

你也可以选择按主题-文档、文档-主题、主题词计数和网络图显示的可视化结果。

主题-文档树形图可视化结果的一个优势是可以让你链接到内容集中的某篇特定文档——便于你阅读这篇文档的原始文献图像或检查OCR文本。

在此你可以点击文档标题，然后点击弹出的白色文本框，然后就将回到**文档浏览器（Doc Explorer）**视图。





Gale数字学术实验室——操作指南

面板视图（续）

在你使用过分析工具后，面板视图的Analyses（分析）模块将会更有意义。

点击此处，用户将会返回到他们利用实验室平台的文本挖掘工具获得的数据分析结果中。

最后，面板视图包含检索历史（Search History）选项卡。在此将列出为创建该内容集所进行的所有检索。只需点击一下，你就可以链接回到检索结果页面，继续完善和更新你的内容集。

Indicators of Global Warming in the Natural World

DOWNLOAD READY

Newspaper articles and editorials relating to global warming.

Overview

Documents (124)

Analyses

Search History

ARCHIVES USED (12)

The Independent Digital Archive (59)
The Times Digital Archive (53)
The Sunday Times Digital Archive (6)
The Telegraph Historical Archive (5)
Daily Mail Historical Archive, 1896-2004 (4)
View More

DOCUMENT TYPE (7)

Article (112)
Essay (5)
Editorial (3)
Letter to the editor (2)
Conference news (1)
View More

AUTHORS (76)

Steve Connor, Science Editor (7)
Lewis Smith (7)
Michael McCarthy, Environment Editor (7)
Michael McCarthy (5)
Nick Nuttall (2)
View More

SOURCE LIBRARIES (12)

The Independent (59)
Times Newspapers Limited (40)
Telegraph Media Group (5)
Associated Newspapers Limited (4)
The New York Times Company (4)
View More

Indicators of Global Warming in the Natural World

DOWNLOAD READY

Newspaper articles and editorials relating to global warming.

Overview

Documents (124)

Analyses

Search History

Clustering

Tabular Data

Scatter Plot

Named Entity Recognition

Hierarchical Tree

Tabular Data

Ngram

Tabular Data

Word Cloud

Bar Chart

Parts of Speech Tagger

Tabular Data

Series View

Overview

Documents (124)

Analyses

Search History

SEARCH	RESULTS	DOCS ADDED	TIMESTAMP
☐ Basic Search ("global warming and climate change")	65	5	Thu Oct 18 09:46:11 EDT 2018
☐ Basic Search (insects And climate change)	10,870	88	Mon Oct 22 09:40:28 EDT 2018
☐ Basic Search (insects And climate change) LIMITS: Publication Date (Jan 01, 1990 - Sep 30, 2018)	831	15	Mon Oct 22 09:43:54 EDT 2018
☐ Basic Search ("effects of global warming") LIMITS: Entire Document (creatures)	26	7	Thu Oct 18 09:50:05 EDT 2018
☐ Basic Search (insects And climate change)	10,870	7	Mon Oct 22 05:55:42 EDT 2018
☐ Basic Search ("effects of global warming")	621	1	Thu Oct 18 09:48:27 EDT 2018

GALE DIGITAL SCHOLAR LAB