



一次史料テキストマイニング・プラットフォーム
Gale Digital Scholar Lab

Galeウェビナー

2024年8月29日

センゲージラーニング株式会社
Gale部門 商品担当

森澤 正樹

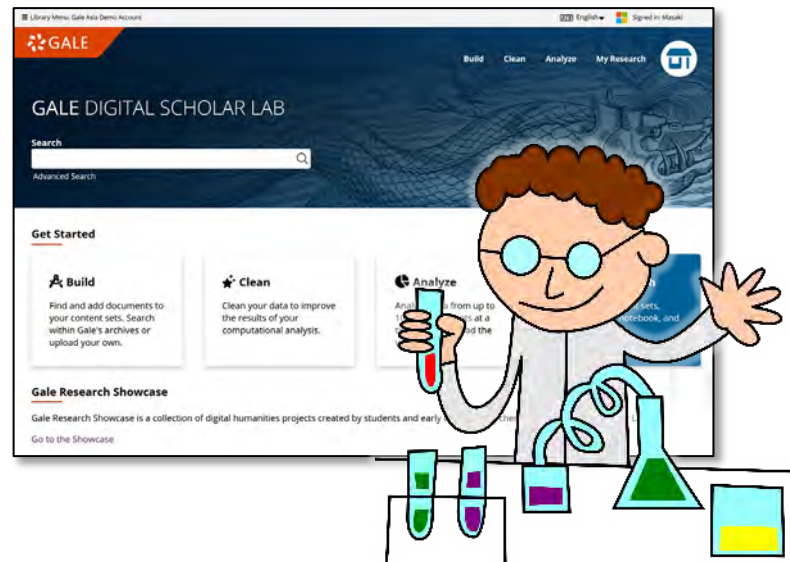
masaki.morisawa@cengage.com

DIGITAL SCHOLAR LAB

Explore your institution's holdings from Gale Primary Sources with new eyes,
using digital humanities to discover new insights on historical materials

本日のアジェンダ

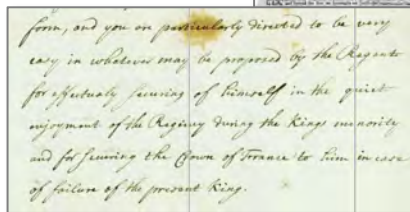
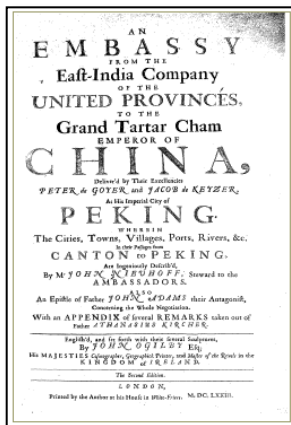
- Gale Primary Sources について
- Gale Digital Scholar Lab について
 - 概要
 - 基本ワークフロー
 - Notebooks/Projects機能
 - Gale Research Showcase
- デモンストレーション
- Q&A



Gale Primary Sources について



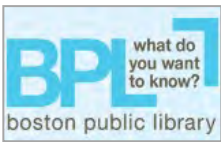
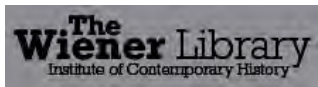
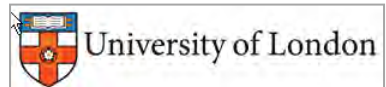
Gale Primary Sources



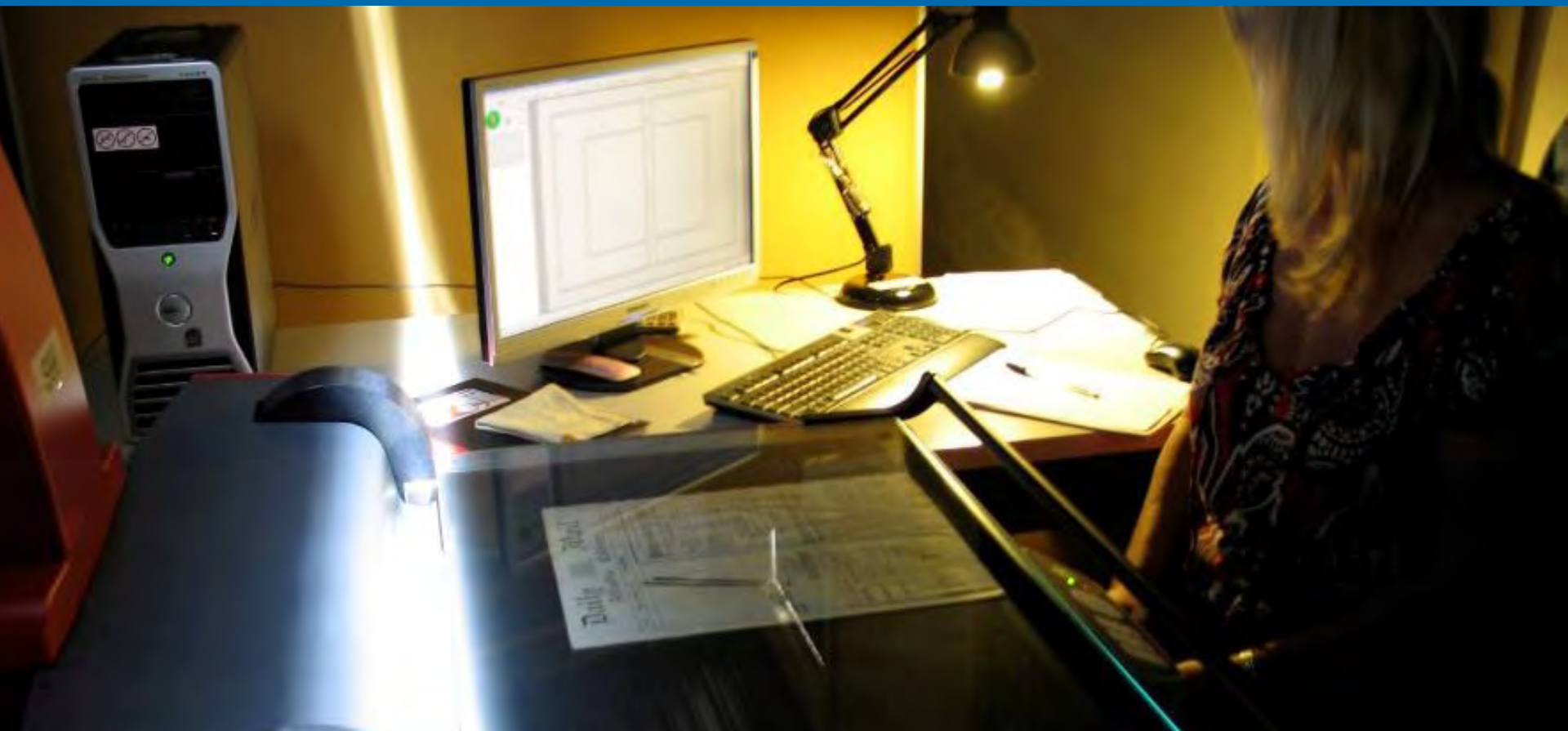
- Gale のデジタル一次史料データベース群の総称
- 同データベース群の横断検索プラットフォームの名称
- データベース数60種以上、総ページ数2億ページ以上
- 主な商品の例:
 - Eighteenth Century Collections Online (ECCO)
 - Nineteenth Century Collections Online (NCCO)
 - The Times Digital Archive
 - The Making of the Modern World (MOMW)
 - U.S. Declassified Documents Online
 - ほか、多数



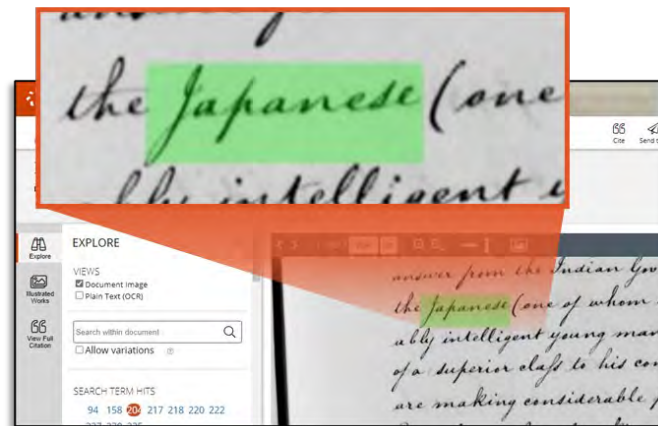
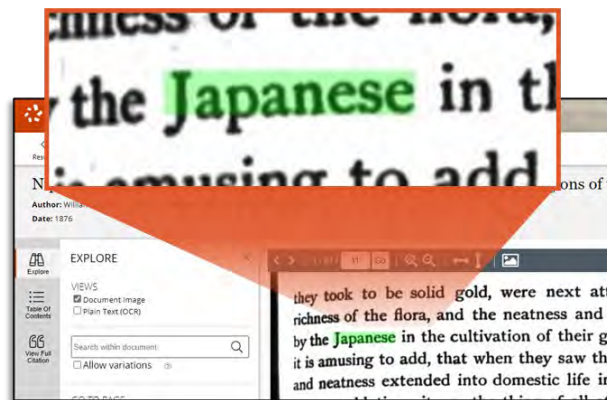
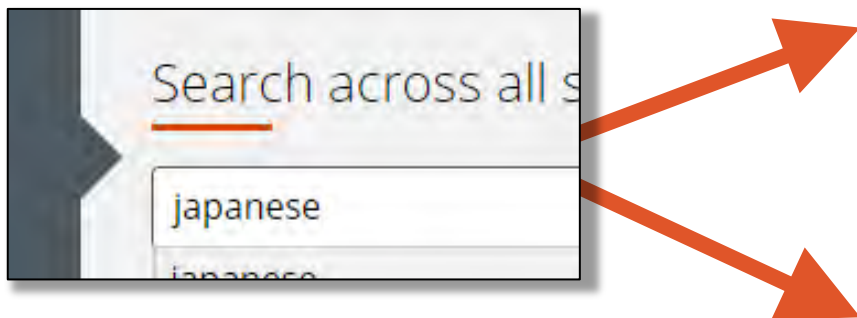
原本収蔵機関の例



貴重資料のデジタル化



歴史資料のフルテキスト検索



OCR = 光学文字認識 (Optical Character Recognition)

HTR = 手書き文字認識 (Handwritten Text Recognition)

* HTRIは一部の最新データベースのみ

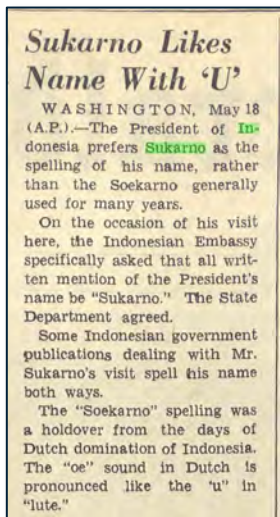
Gale Digital Scholar Lab について



GALE DIGITAL SCHOLAR LAB

EMPOWER™ RESEARCH

文書の背後のOCRテキスト



文書画像

Sukarno Likes Name with 'U' Sukarno Likes Name With 'U', WASHINGTON, May 18 (A.P.). The President of Indonesia prefers Sukarno as the spelling of his name, rather than the Soekarno generally used for many years. On the occasion of his visit here, the Indonesian Embassy specifically asked that all written mention of the President's name be "Sukarno." The State Department agreed. Some Indonesian government publications dealing with Mr. Sukarno's visit spell his name both ways. The "Soekarno" spelling was a holdover from the days of Dutch domination of Indonesia. The "oe" sound in Dutch is pronounced like the "u" in "lute."

OCRテキスト



OCRテキストをたくさん集めて、分析できたら？

Sukarno Likes Name with 'U' Sukarno Likes Name With 'U' WASHINGTON, May 18 (A.P.) The President of Indonesia prefers Sukarno as the spelling of his name, rather than the Soekarno generally used for many years. On the occasion of his visit here, the Indonesian Embassy specifically asked that all written mention of the President's name be 'Sukarno.' His State Department agreed. Some Indonesian government publications dealing with Mr. Sukarno's visit spell his name both ways. The 'Soekarno' spelling was a holdover from the days of Dutch domination of Indonesia. The 'oe' sound in Dutch is pronounced like the 'ou' in 'oute.'



Eisenhower, Sukarno Agree on 'Freedom' 881 mm President Sukarno of Indonesia, and Mrs. Richard M. Nixon at a Washington reception. Eisenhower, Sukarno see on 'Freedom' Ag. By the United Press WASHINGTON, May 20. President Eisenhower said Friday night that the people of the United States must never forget that unless we defend the freedom of the other fellow we are endangering our own. Mr. Eisenhower, in a toast to visiting Indonesian President Sukarno, said that he wanted to express his gratitude to Mr. Sukarno for recalling a basic American responsibility. 'I want to thank you for what you have reminded us of.'



Sukarno Hailed on Broadway Sukarno Hailed On Broadway. NEW YORK, May 23 (A.P.) New York today accorded President Sukarno of Indonesia the traditional honor for visiting dignitaries a ticker tape parade along Lower Broadway through the skyscrapers of the financial district. The President stood in the rear of an open car for the 15-minute drive from Battery Park to City Hall. A light rain began to fall just before the parade started, but even so police estimated the crowd along the line of march and in City Hall Plaza at 80,000. Many persons along the parade route waved the red and white flag of Indonesia. Mr. Sukarno waved repeatedly. Arriving at City Hall, Mr. Sukarno was greeted by Mayor Robert F. Wagner. Both men stood at attention as the national anthems of the United States and Indonesia were played.



Sukarno and Pope Hold Talk Sukarno And Pope Hold Talk By Barrett McGurn From the Herald Tribune Bureau (g) 1956 New York Herald Tribune, Inc.ROME, June 13. Pope Plus XII promised President Achmed Sukarno of Indonesia here today that foreign missionaries of the Roman Catholic Church will be withdrawn from his country as quickly as possible, and appealed in effect for a transition period. The Indonesian revolutionary nationalist leader assured the Pope in reply that he considered the 1,000,000 Indonesian Catholics, slightly more than 1 per cent of the population, a loyal part of his country and wished the Catholic religion well in his homeland.



...

=



2つの異なるワークフロー

GALE PRIMARY SOURCES

thailand

Advanced Search

West of the Mekong

Publication: The Economist

GALE PRIMARY SOURCES

Search

Advanced Search

Clear Save Download Print Get Link

West of the Mekong

Publication: The Economist

Table of Contents

View Full Article

West of the Mekong

Publication: The Economist

GALE PRIMARY SOURCES

検索 → 閲読

Clustering of documents in your collection.

Cluster-1
Cluster-2
Cluster-3
Cluster-4
Cluster-5
Cluster-6

Topic Proportion by Document

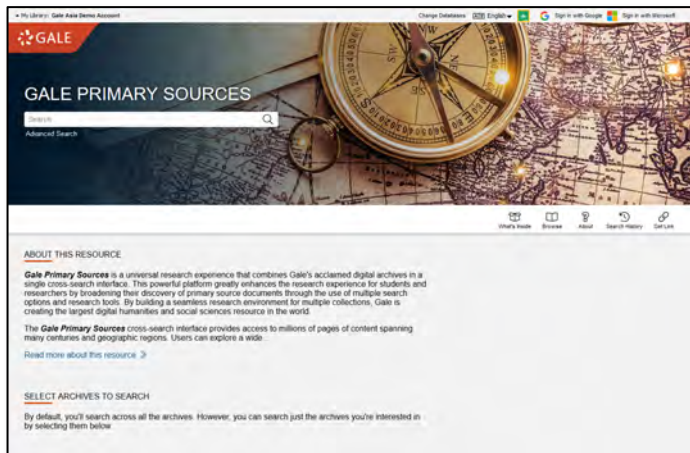
Barney in Greece
The Diesel Engine
Freight
Hunan Takes Aim
The Loss of Bataan
Multiple News Items
Wants to Amok
Local Revivals
Wyoming Man Selected
Treaty Signed to Annex Hawaii

DOCUMENTS DISPLAYED (15):

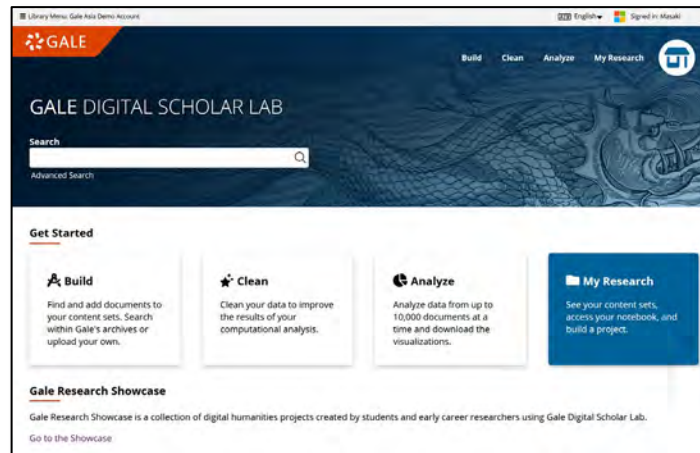
Selected Documents

データ収集 → 分析

それぞれに最適なプラットフォーム



Gale Primary Sources



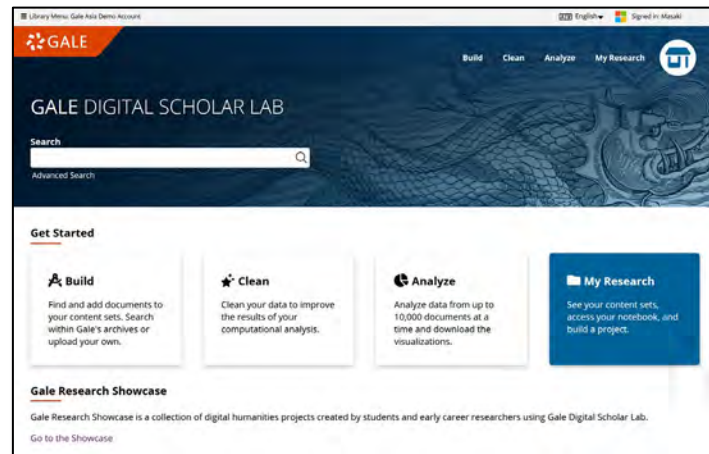
Gale Digital Scholar Lab



Gale Digital Scholar Lab

Galeのテキストマイニング専用プラットフォーム

- プログラミング等の知識は不要！
- Gale 一次史料データベース群のOCR/HTRテキストを搭載
- テキスト分析ツール／OCRクリーニング・ツールを内蔵
- グループワークに対応
- 分析結果についてのエッセイの執筆が可能《新機能》
- 発表の場「Gale Research Showcase」を新設《新機能》



Gale Digital Scholar Lab の基本ワークフロー



Gale Digital Scholar Lab の基本ワークフロー



現在 Lab で分析可能なデータベース

- Amateur Newspapers from the American Antiquarian Society
- American Fiction 1774-1920
- American Historical Periodicals from the American Antiquarian Society
- Archives Unbound
- Archives of Latin American and Caribbean History
- Archives of Sexuality and Gender
- Associated Press Collections Online
- Brazilian and Portuguese History and Culture
- British Library Newspapers
- Chatham House Online Archive
- China and the Modern World
- Crime, Punishment, and Popular Culture, 1790-1920
- Daily Mail Historical Archive 1896-2016
- Declassified Documents Online: Twentieth Century British Intelligence
- Decolonization
- The Economist Historical Archive, 1843-2020
- Eighteenth Century Collections Online (ECCO)
- Environmental History
- History of Disabilities
- The Illustrated London News Historical Archive 1842-2003
- The Independent Historical Archive
- Indigenous Peoples of North America
- International Herald Tribune Historical Archive, 1887-2013
- Liberty Magazine Historical Archive, 1924-1950
- The Listener Historical Archive, 1929-1991
- Mail on Sunday Historical Archive, 1982-2011
- The Making of Modern Law (MOML)
- The Making of the Modern World (MOMW)
- Mirror Historical Archive 1903-2000
- Nineteenth Century Collections Online (NCCO)
- Nineteenth Century U.S. Newspapers
- Nineteenth Century UK Periodicals
- Picture Post Historical Archive 1938-1957
- Political Extremism and Radicalism
- Power to the People
- Public Health Archives: Public Health in Modern America, 1890-1970
- Punch Historical Archive 1841-1992
- Refugees, Relief, and Resettlement
- Religions of America
- Sabin Americana: History of the Americas, 1500-1926
- Seventeenth and Eighteenth Century Burney Newspapers Collection
- Seventeenth and Eighteenth Century Nichols Newspapers Collection
- Slavery and Anti-Slavery: A Transnational Archive
- Smithsonian Collections Online
- State Papers Online Colonial
- The Sunday Times Historical Archive 1822-2016
- The Telegraph Historical Archive 1855-2016
- The Times Digital Archive 1785-2019
- The Times Educational Supplement Historical Archive, 1910–2000
- The Times Literary Supplement Historical Archive 1902-2019
- U.S. Declassified Documents Online
- Women's Studies Archive

現在 Lab で分析できないデータベース

- Financial Times Historical Archive
- National Geographic Magazine Archive
- Early Arabic Printed Books from the British Library
- British Literary Manuscripts Online
- State Papers Online: Early Modern Government in Britain and Europe*

* 似た名称の「State Papers Online Colonial」は分析可能

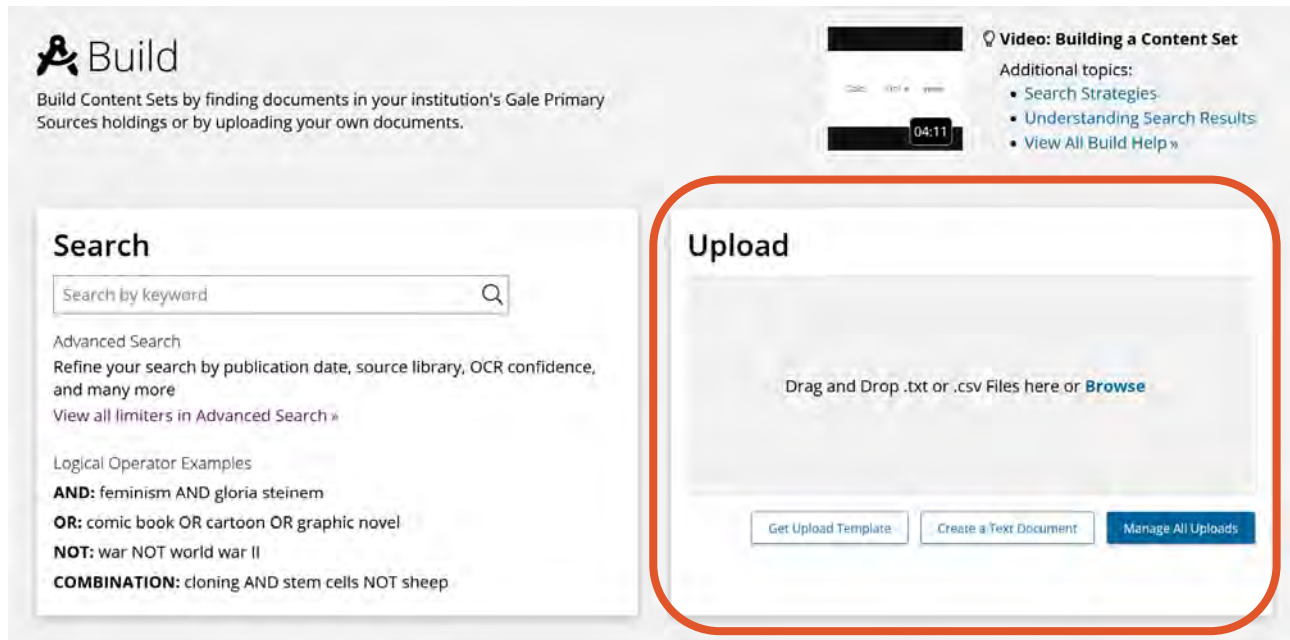
主な理由:

- 権利関係の理由 (Financial Times, National Geographic)
- 言語の問題 (Early Arabic Printed Books)
- 手稿中心でOCRがない (British Literary Manuscripts、State Papers Online)



自身のコンテンツをアップロードする

- テキストを直接入力するか、TXT形式／CSV形式のファイルをアップロードする
- 1度にアップロードできるファイルは最大10MBまで
- 著作権・テキストマイニング権のクリアランスはユーザー自身の責任となります。



The screenshot displays the Gale Build interface. On the left, the 'Search' section includes a 'Search by keyword' input field, an 'Advanced Search' section with a description and a link to 'View all limiters in Advanced Search', and 'Logical Operator Examples' such as 'AND: feminism AND gloria steinem', 'OR: comic book OR cartoon OR graphic novel', 'NOT: war NOT world war II', and 'COMBINATION: cloning AND stem cells NOT sheep'. On the right, the 'Upload' section is highlighted with a red rounded rectangle and contains the text 'Drag and Drop .txt or .csv Files here or Browse' and three buttons: 'Get Upload Template', 'Create a Text Document', and 'Manage All Uploads'. A video player in the top right corner shows a video titled 'Video: Building a Content Set' with a duration of 04:11 and a list of 'Additional topics' including 'Search Strategies', 'Understanding Search Results', and 'View All Build Help'.



Gale Digital Scholar Lab の基本ワークフロー



OCRクリーニング・ツールの設定(任意)

- OCRのクリーニング項目を設定することができる
- ストップワード*、大文字・小文字の区別、数字・句読点・特殊文字の扱い、語句の検索置換など
- 複数のクリーニング設定を作成・保存し、分析ごとに適用することができる

* ストップワード(stop words) : あまりに一般的であるため、分析の対象から除外する語。英語の場合、「a」「the」「of」「in」等々。



Library Menu GALE Asia Demos Account

GALE DIGITAL SCHOLAR LAB

Build Clean Analyze My Research

Add Note New Configuration Save All Test Configuration

Clean

Select Cleaning Configuration to Edit
Default Cleaning Configuration

Video: Cleaning a Content Set
Additional topics:
• Creating a Clean Configuration
• Applying During Analysis
• View All Clean Help

Stop words

Set the words you want the Analysis Tools to ignore.
Choose a Starter List

- a
- about
- above
- across
- after
- afterwards
- again
- against
- all
- almost
- alone
- along
- already
- also
- although
- always
- am
- among
- amongst
- amongst
- amount
- an
- and
- another
- any
- anyhow
- anyone
- anything
- anyway
- anywhere
- are
- around
- as
- at
- back
- be
- became

ignore stop words case

Clean List

Text Correction

Options for automatic text correction that will be applied before each Analysis

- Turn on all options
- Text Modification**
 - All lower case
- Characters**
 - Remove all extended ASCII characters
 - Remove all number characters
- Special Characters**
 - Remove all special characters
 - Set specific characters >
- Punctuation**
 - Remove all punctuation
 - Set specific punctuation >
- Spacing**
 - Remove all tabs
 - Remove all line breaks
 - Reduce multiple spaces to one space (ex: "hello there" becomes "hello there")
- Document Sections**
 - Remove body text
 - Remove all non-body content
 - Set specific sections >
- Replacements**

Replace this...	With this...
<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>

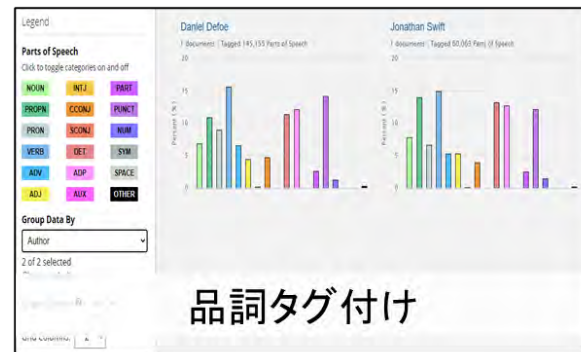
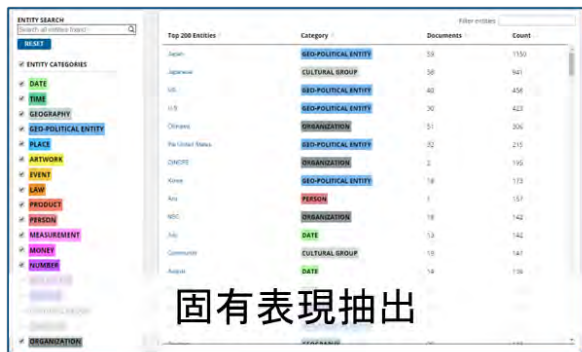
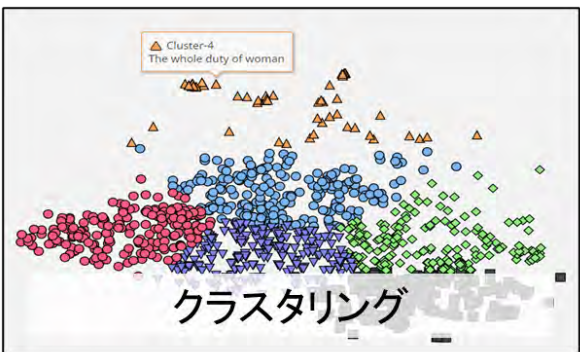
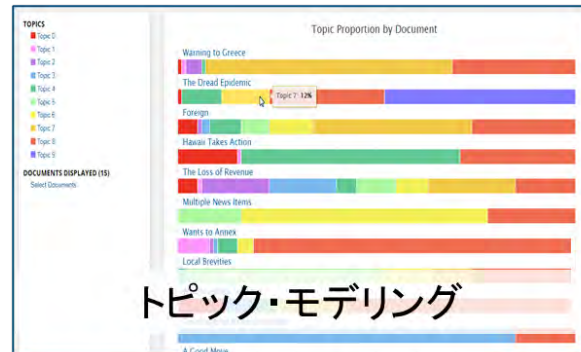
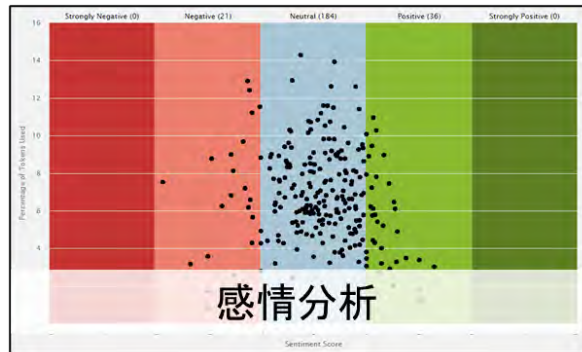
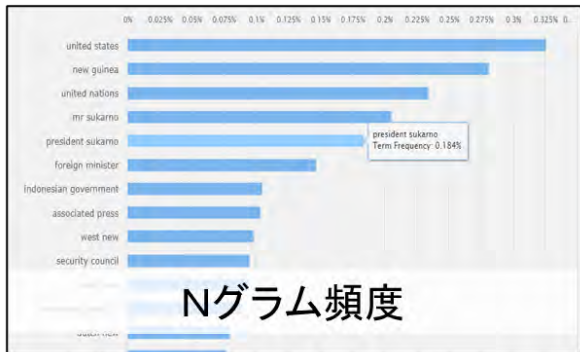
Add a Row



Gale Digital Scholar Lab の基本ワークフロー



Gale Digital Scholar Lab で現在利用可能な分析ツール



ノートブック機能 (Note)

- Labの随所で、気づいた点や疑問点、留意事項などのメモ (Note) を作成できます。
- 分析結果ページで作成したNoteには分析画像を添付することができます。
- グループアカウントで作成されたNoteはメンバー全員が閲覧できます。
- 作成したNoteはProject (後述) に割り当てることができます。

The screenshot displays the GALE DIGITAL SCHOLAR LAB interface. The main window shows a Topic Modeling analysis for "Japanese in TTDA 1904-1905". The interface includes a Legend, Topic Metric (Tokens), Sort (Metric High-Low), and Topics (Battle details, Japanese politics, Soldiers in trenches, Japanese military, British politics, Naval battle, Treaty negotiations). A bar chart shows the distribution of tokens across topics. An "Add Note" dialog box is open, showing the text: "The 'Japanese military' topic dominates the token count, followed closely by 'Correspondents', showing the importance of wartime correspondents, both of the Times and of other papers, in reporting the conflict. Interestingly there is no clearly discernible topic on the Russian side of the conflict, perhaps revealing the paper's bias towards Japan." The dialog box has a "NOTE" title, a rich text editor, and options: "Attach Visualization" (checked), "Keep my name anonymous" (unchecked), and "See all notes in Notebook". A red box highlights the "Attach Visualization" option, with an arrow pointing to the text "分析結果のビジュアルを添付する".

分析結果のビジュアルを添付する



Attach Visualization

Keep my name anonymous

See all notes in Notebook

Cancel

Save

エッセイを執筆する(Projects機能)

- Lab内でエッセイを執筆できる機能です。
- Projectに割り当てたNoteを挿入して編集したり、新たにヘッダーや本文を作成し、外部画像やリンク等を挿入して並べ替えることができます。
- ProjectはPDF形式でダウンロードできるほか、Galeに提出し、簡単な査読を経て、Gale Research Showcase (後述)で公開することもできます。

Note機能で作成したメモ

NOTE

Tokens by Token Count

The "japanese military" topic dominates the token count, followed closely by "Correspondents", showing the importance of wartime correspondents, both of the Times and of other papers, in reporting the conflict. Interestingly there is no clearly discernible topic on the Russian side of the conflict, perhaps revealing the paper's bias towards Japan.

Attach Visualization
 Keep my name anonymous

See all notes in Notebook

Cancel Save

直接書き下ろしたコンテンツ

EDIT

Section Subheading (Optional)
Introduction

TSO characters allowed

Body Text

The Russo-Japanese War of 1904-1905 was a significant conflict that heralded the rise of Japan as a major world power and marked the decline of Tsarist Russia. This war, fought mainly over imperial ambitions in Manchuria and Korea, captivated global audiences, including those in Great Britain. The *London Times*, one of the most influential newspapers of the time, covered the war extensively. Analyzing its content through text mining methods can provide rich insights into the narratives, biases, and attitudes conveyed to the British public during the conflict.

Cancel Save

GALE DIGITAL SCHOLAR LAB

Build | Data | Analyze | My Research

Text Mining the London Times During the Russo-Japanese War of 1904-1905

This study examines how the London Times informed and shaped public opinion during the Russo-Japanese War of 1904-1905 by examining recurring vocabulary, topics, and sentiments found in the OCR text.

Introduction

The Russo-Japanese War of 1904-1905 was a significant conflict that heralded the rise of Japan as a major world power and marked the decline of Tsarist Russia. This war, fought mainly over imperial ambitions in Manchuria and Korea, captivated global audiences, including those in Great Britain. The *London Times*, one of the most influential newspapers of the time, covered the war extensively. Analyzing its content through text mining methods can provide rich insights into the narratives, biases, and attitudes conveyed to the British public during the conflict.

Methodology

The text mining approach in this study revolves around three core tasks: vocabulary extraction (Ngrams), topic modeling, and sentiment analysis. The OCR text from the London Times papers published in the years 1904 to 1905 provides the primary dataset.

Ngrams Analysis

Topic Modeling

Topics by Token Count

The "japanese military" topic dominates the token count, followed closely by "Correspondents", showing the importance of wartime correspondents, both of the Times and of other papers, in reporting the conflict. Interestingly there is no clearly discernible topic on the Russian side of the conflict, perhaps revealing the paper's bias towards Japan.

Tokens

Token	Count
the	1000
and	800
of	700
in	600
to	500
a	400
was	300
an	200
on	150
at	100
from	80
with	70
by	60
and	50
of	40
the	30
in	20
to	15
a	10
was	8
an	6
on	4
at	3
from	2
with	1
by	1

さらに、出来上がったエッセイをGaleに提出して...



Close Preview

Text Mining the London Times During the Russo-Japanese War of 1904-1905

Introduction

Methodology

Ngrams Analysis

Topic Modeling

Sentiment Analysis

Conclusion

Text Mining the London Times During the Russo-Japanese War of 1904-1905

This essay examines how the London Times informed and shaped public opinion during the Russo-Japanese War of 1904-1905 by examining recurring vocabulary, topics, and sentiments found in the OCR text.

Introduction

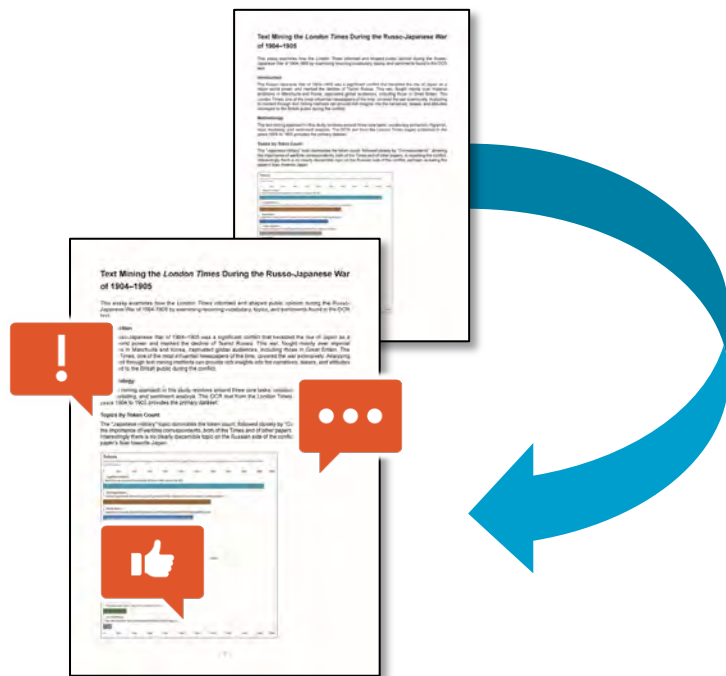
The Russo-Japanese War of 1904-1905 was a significant conflict that heralded the rise of Japan as a major world power and marked the decline of Tsarist Russia. This war, fought mainly over imperial ambitions in Manchuria and Korea, captivated global audiences, including those in Great Britain. The London Times, one of the most influential newspapers of the time, covered the war extensively. Analyzing its content through text mining methods can provide rich insights into the narratives, biases, and attitudes conveyed to the British public during the conflict.

Methodology

Ngrams Analysis

The text mining approach in this study revolves around three core tasks: vocabulary extraction (Ngrams), topic modeling,

デジタル人文学研究者による査読をクリアすると...



(査読者はイメージです)

Gale Research Showcase で広くオープンに公開されます！



www.gale.com/dh-research

20世紀研究

(このエッセイははめ込み合成です)

Masaki Morisawa (University of Gale) | [Text Mining the London Times During the Russo-Japanese War of 1904–1905](#)

Aryan Shah, Livia Ngo, Kody Chantavong and Megan Skrobot (University of Washington) | [Roberto Calvi's Trial: Suicide or Murder?](#)

[Black America and the Law in the Mid-20th Century](#)

[The Rise of Electricity in the Late 19th and Early 20th Centuries](#)

[Tracking Archaeology in The Illustrated London News](#)

学生が切り開くテキスト

Gale Digital Scholar Lab は、プログラミング技術の有無に関係なく、あらゆる種類のデータ分析の経験は、新たな洞察に結び付くだけでなく、デジタルデータに集められたプロジェクトは、様々な学生の方が Lab を使用して創造性を発揮する例が、Gale Digital Scholar Lab で新たなプロジェクトを始めるためのヒントとご紹介します。新たなプロジェクトは Gale Digital Scholar Lab 内の Projects機能を用いて提出することができます。

分野別プロジェクト一覧

17世紀研究

18世紀研究

19世紀研究

20世紀研究

21世紀研究

EMPOWER™ RESEARCH

デモンストレーション



ご清聴ありがとうございました。

無料トライアルのお申込み・お問い合わせは下記まで

GaleJapan@cengage.com

www.gale.com/jp

 x.com/gale_japan