



デジタル人文学〈3つの障害〉 と Gale Digital Scholar Lab

2020年1月25日

Galeシンポジウム2020

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

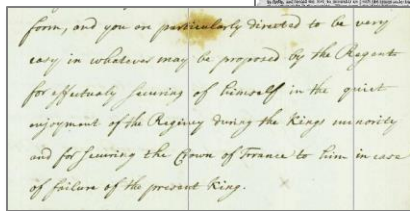
会社紹介



- 米国の教育出版社センゲージラーニングの図書館部門
- 本部: 米国ミシガン州ファーミントン・ヒルズ市
- 出版形態
 - ✓ 百科事典・年鑑などのレファレンス書籍
 - ✓ 電子ブック・文学情報・雑誌記事などのデータベース
 - ✓ 一次資料・新聞などのマイクロフォーム
 - ✓ 一次資料・新聞などのデジタル・アーカイブ

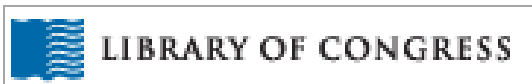


Gale Primary Sources



- 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300種以上、総ページ数170万ページ以上
- 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



LIBRARY OF CONGRESS

AMERICAN ANTIQUARIAN SOCIETY



University of London



NATIONAL LIBRARY of IRELAND



The Wiener Library
Institute of Contemporary History



THE CHURCH OF ENGLAND



UNIVERSITY OF OXFORD



BRITISH LIBRARY



Yale University



THE HUNTINGTON
Library, Art Collections, and Botanical Gardens



MANCHESTER
1824



The National Archives



The New York Public Library



nls National Library of Scotland



OBERLIN
COLLEGE & CONSERVATORY



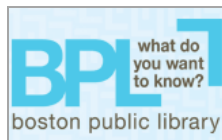
THE SALVATION ARMY



Tulane University



HARVARD UNIVERSITY



BPL what do you want to know?
boston public library



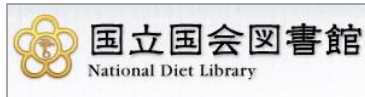
NATIONAL ARCHIVES



NATIONAL LIBRARY OF AUSTRALIA



UNIVERSITY OF CAMBRIDGE



国立国会図書館
National Diet Library



國家圖書館出版社
National Library of China Publishing House
任地念題



国立大学法人
長崎大学 附属図書館
Nagasaki University Library

貴重資料のデジタル化



OCR (光学文字認識)によるフルテキスト検索

GALE PRIMARY SOURCES

osaka

Advanced Search

Justice: The
kyo, **Osaka**, Send
Courts and Dist

na-Pacific International Exhibition, 1915: Prepared by Hakurankwai

which exercises the high function of council to the Emperor.
Justice is administered by the Courts of Justice in the name of the Emperor.

There are four kinds of Courts of Justice: The Supreme Court (in Tokyo), Courts of Appeal (in Tokyo, **Osaka**, Sendai, Hiroshima, Nagoya, Nagasaki, Hakodate), Local Courts, and District Courts. In these four kinds of Courts of Justice, civil as well as criminal judgments are awarded. In Formosa, there are two kinds of Courts of Justice: Courts of Appeal and Local Courts. In Korea, there are four kinds of courts: The Higher Court, the Court of Appeal, the Local Courts, and the District Courts.

b. "FU" and "KEN" (Prefecture)

The Empire of Japan is divided into three "Fu" and forty three "Ken," as administrative districts. Besides, there are established "Cho" in Hokkaido and Saghalien; and governments-

View full citation

Allow variations

Relevant Pages

18	31	51	59	60	87	
149	150	151	152	154	158	160
161	163					

Related Subjects

OCR = Optical Character Recognition

Gale Primary Sources 商品群



...ところが、

『タイムズ』紙と婦人参政権運動



TO THE EDITOR OF THE TIMES.

Sir,—May I express my entire agreement with the letter of Miss Milner in your issue of this morning? If the recent scenes of rowdiness associated with women's franchise only served to bring ridicule on the self-appointed champions of that cause other women might be well content to let the matter rest there. Unfortunately, such behaviour can only have the most mischievous effect in prejudicing the influence of women in those branches of public life where the beneficial character of their work is universally recognized.

It is often said of women that neither logic nor humour counts among their strongest points. The recent behaviour of the **suffragettes** would appear to support this contention. Mrs. Fenwick Miller's letter in *The Times* this morning is in every way a remarkable document. It opens up an attractive vista of the public results we might expect to follow from the establishment of feminine rule marked by such a judicious and temperate spirit, say, at the Board of Trade or India Office. As an onlooker nothing strikes me as more curious in this controversy than the unreasonable but most feminine desire of the **suffragettes** both to eat and to keep their political and domestic cake. Women cannot expect to have it both ways. They cannot at one and the same time

1913_ML.mut

N	Word 1	Freq	Word 2	Freq	Texts	Gap	Join	M	Z
1	SUFFRAGISTS	615	ATTRIBUTED	56	20	2	22	9.83	44.29
2	SUFFRAGISTS	527	COLOURS	39	10	1	11	9.58	28.61
3	SUFFRAGISTS	615	MILITANT	353	93	1	116	9.58	92.82
4	SUFFRAGIST	527	BOMBS	25	4	1	7	9.57	22.73
5	SUFFRAGE	1,386	WVOMAN	30	18	1	22	9.56	40.21
6	SUFFRAGIST	527	LITERATURE	87	21	1	23	9.48	40.01
7	SUFFRAGIST	527	BREAKERS	18	3	2	5	9.56	19.14
8	SUFFRAGE	1,386	WOMAN	1,178	253	1	683	9.22	198.56
9	SUFFRAGE	1,386	ADULT	63	25	1	43	9.46	54.17
10	SUFFRAGISTS	615	ANTI	168	30	1	43	9.21	49.69
11	SUFFRAGIST	527	ANTI	168	23	1	33	9.06	41.11
12	SUFFRAGE	1,386	WOMAN	25	9	1	12	8.95	23.85
13	SUFFRAGE	1,386	OPPOSING	66	22	2	30	8.87	36.66
14	SUFFRAGE	1,386	MANHOOD	31	13	1	16	9.05	28.60
15	SUFFRAGIST	527	PUBLICATIONS	35	4	1	6	8.86	16.33
16	SUFFRAGIST	527	OUTRAGE	128	19	1	21	8.80	29.86
17	SUFFRAGISTS	615	SUSPECTED	43	8	2	8	8.75	18.15
18	SUFFRAGIST	527	DISORDER	92	9	3	13	8.58	21.72
19	SUFFRAGIST	527	CONSPIRACY	90	11	1	13	8.61	21.97
20	SUFFRAGE	1,386	SOCIETIES	195	48	1	72	8.57	50.93
21	SUFFRAGE	1,386	FEMALE	113	23	1	41	8.55	38.08
22	SUFFRAGE	1,386	FEDERATED	14	5	3	5	8.52	13.19
23	SUFFRAGIST	527	SUPPOSED	132	16	1	17	8.45	23.65
24	SUFFRAGIST	527	DISTURBANCE	82	9	1	11	8.51	19.44

frequency alphabetical statistics filenames mutual information notes

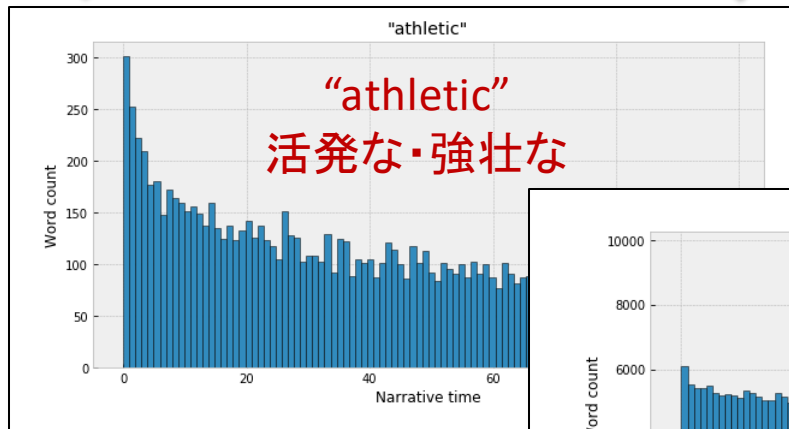
330 Type-in OUTRAGE

初期アメリカ小説と用語頻度

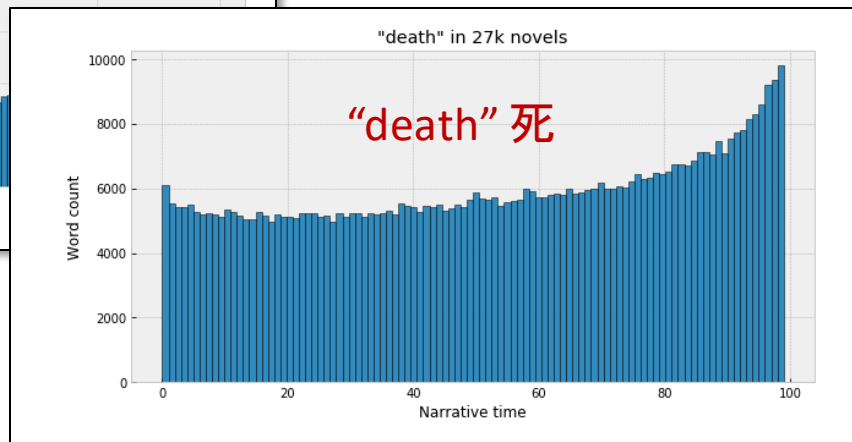
GALE
AMERICAN
FICTION
1774-1920



小説の冒頭



小説の末尾

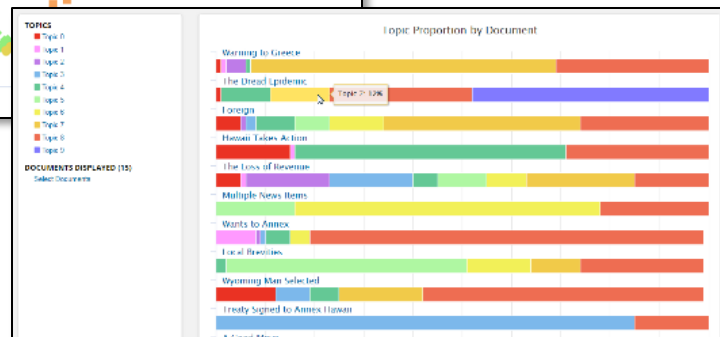
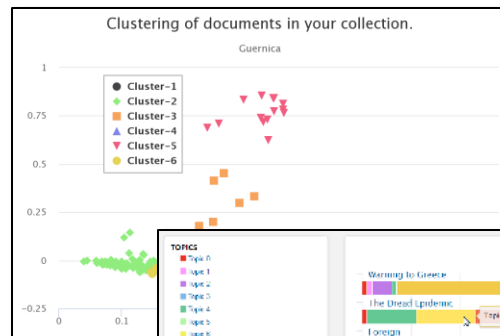


スタンフォード文学ラボ デビッド・マクルーア先生

<https://litlab.stanford.edu/distributions-of-words-27k-novels/>

2つの異なるワークフロー

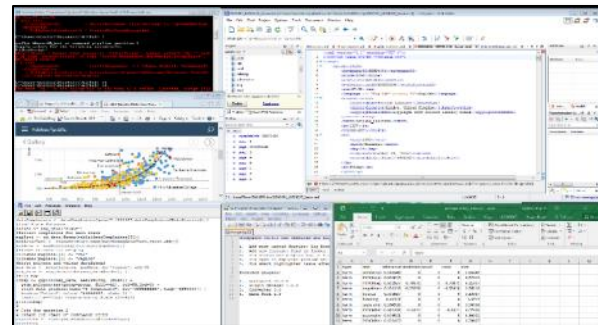
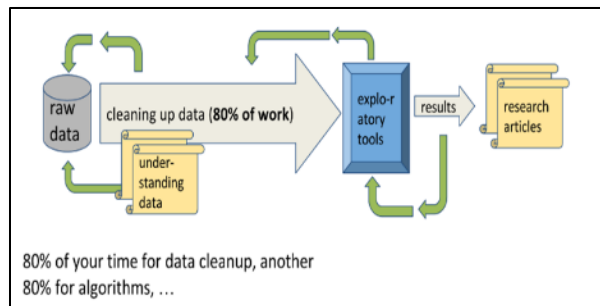
The screenshot shows the Gale Primary Sources interface. At the top, there's a search bar with 'thailand' entered. Below it, the 'Advanced Search' button is visible. The main content area displays a document titled 'West of the Mekong' from 'The Economist'. The document text discusses Thailand's political and economic situation, mentioning the 'Diplomatic Dictionary' and 'Bangkok'. A map of Thailand is also shown on the right side of the document page.



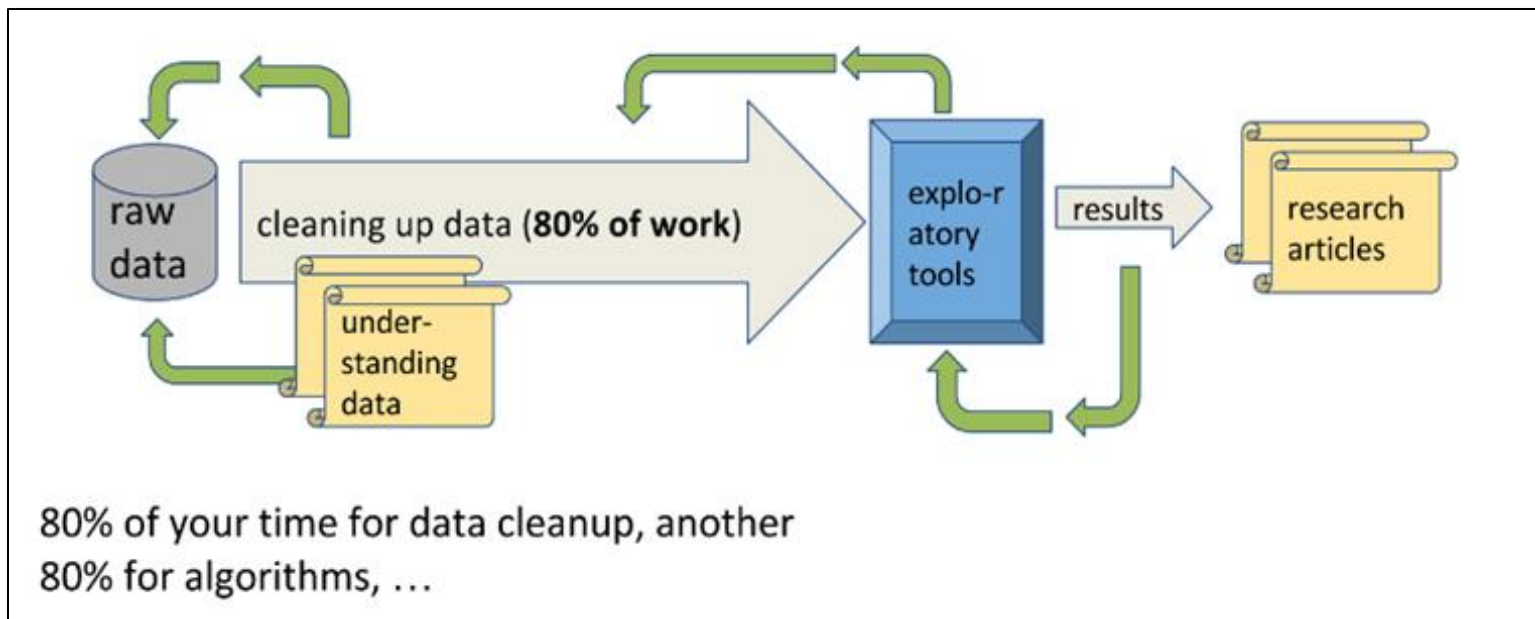
検索 → 閲読

データ収集 → 分析

デジタル人文学：3つの障害



1: 分析用に最適化された元データの入手



Slide courtesy of the COMHIS Collective: <https://comhis.github.io/> <http://j.mp/comhis-bsecs>

2: データの保存場所の確保と管理

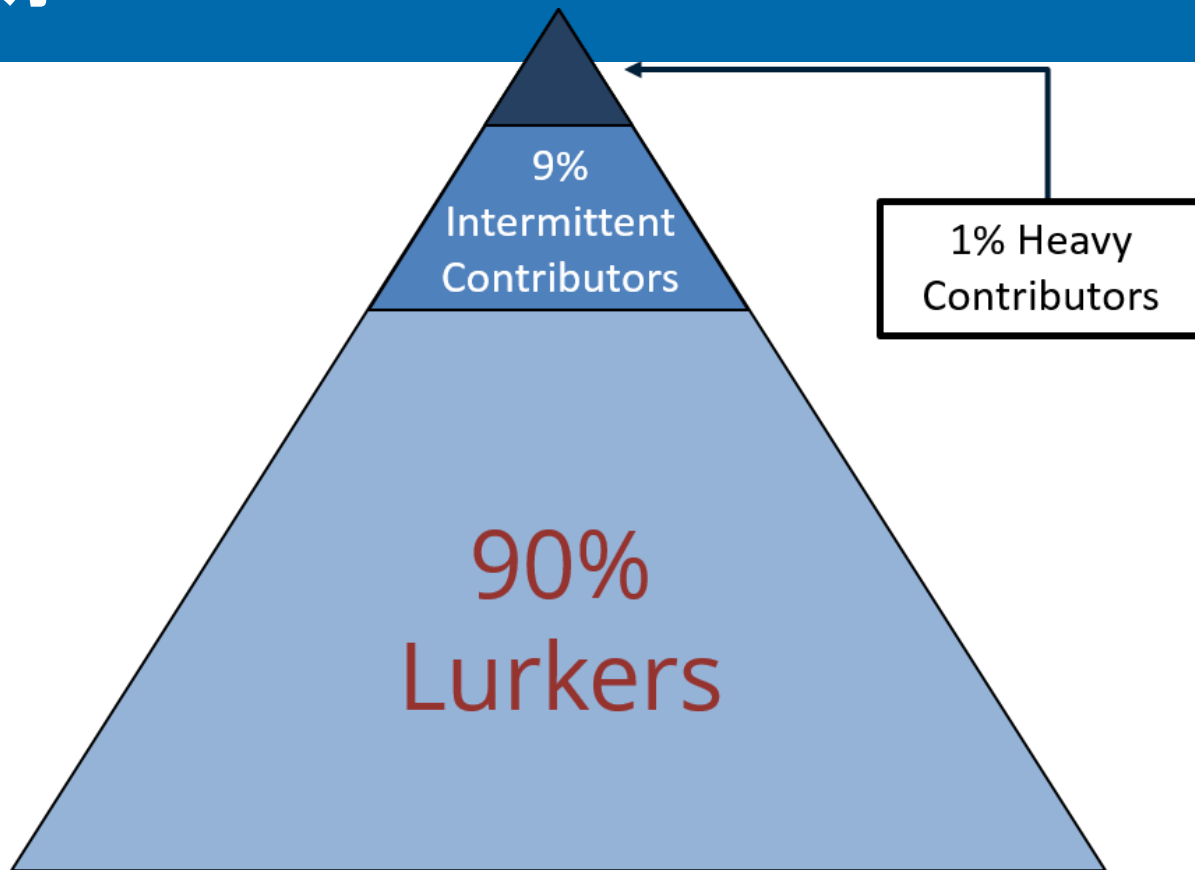


3: 分析ツールの難しさ

The collage illustrates the complexity of data analysis tools through five different screenshots:

- Terminal/Script:** Shows a complex shell script with various commands, variables, and conditional logic, representing the complexity of command-line tools.
- File Explorer:** Shows a directory structure with subfolders like 'input', 'output', and 'logs', representing the complexity of file management and organization.
- Code Editor:** Shows a complex JSON or XML configuration file with many nested elements, representing the complexity of configuration management.
- Scatter Plot:** Shows a scatter plot visualization with data points in various colors and a regression line, representing the complexity of data visualization and analysis.
- Spreadsheet:** Shows a table with columns for 'Type', 'Unit', 'Value', 'Count', 'Color', and 'Size', containing numerical data, representing the complexity of data storage and manipulation.

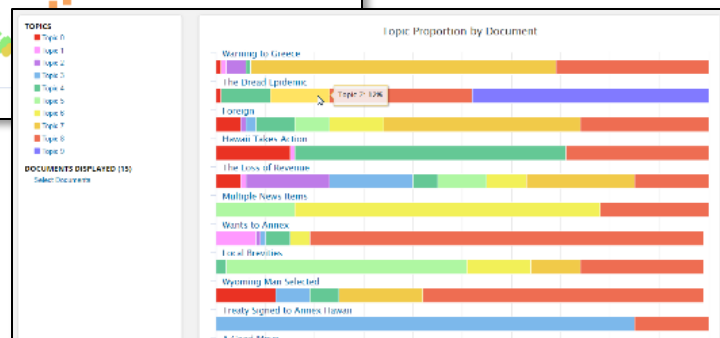
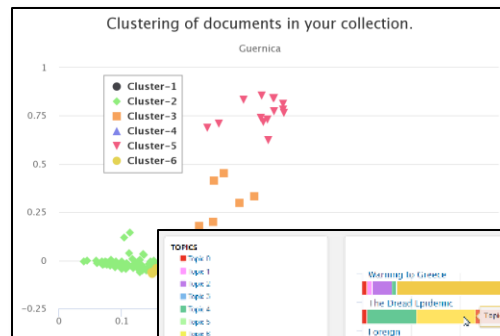
‘1-9-90’ の法則



The 90-9-1 rule for participation in an online community
<http://www.nngroup.com/articles/participation-inequality/>

2つの異なるワークフロー

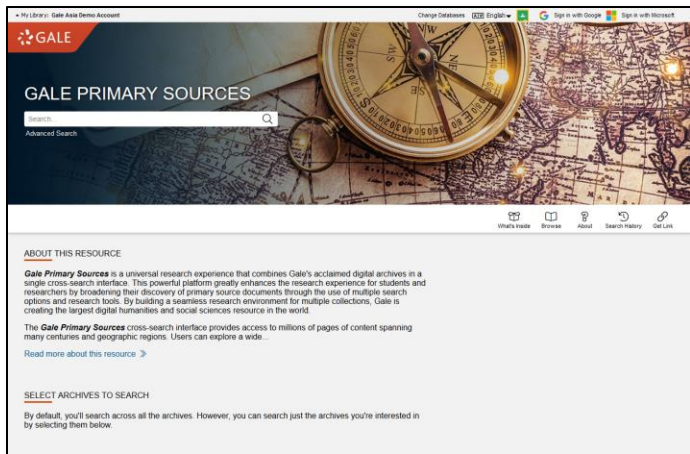
The screenshot shows the Gale Primary Sources interface. At the top, the search term 'thailand' is entered in the search bar. Below the search bar, the results for 'West of the Mekong' are displayed. The article is from 'The Economist' and includes a map of Thailand. The text discusses the political and economic situation in Thailand at the time, mentioning the 'Kingdom of Siam' and the 'Kingdom of Laos'.



検索 → 閲読

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

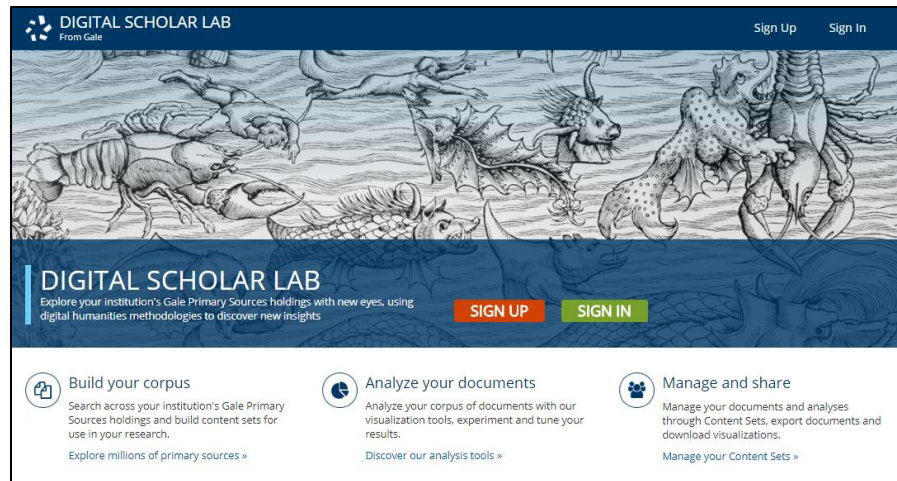


Gale Digital Scholar Lab

Gale Digital Scholar Lab

Galeの一次史料テキストマイニング専用プラットフォーム

- Gale Primary Sources データベース群のOCRテキストを搭載
- 様々な検索条件指定で分析対象となるコンテンツのサブセットを自由に作成できる
- 主要なテキストマイニング・ツールを搭載
- OCRクリーニング・ツールを内蔵
- 分析結果はエクスポート可能(画像・表形式)
- OCRテキストは文書5000点まで一度にエクスポート可能



DIGITAL SCHOLAR LAB
From Gale

Sign Up Sign In

DIGITAL SCHOLAR LAB
Explore your institution's Gale Primary Sources holdings with new eyes, using digital humanities methodologies to discover new insights.

SIGN UP SIGN IN

Build your corpus
Search across your institution's Gale Primary Sources holdings and build content sets for use in your research.
Explore millions of primary sources >

Analyze your documents
Analyze your corpus of documents with our visualization tools, experiment and tune your results.
Discover our analysis tools >

Manage and share
Manage your documents and analyses through Content Sets, export documents and download visualizations.
Manage your Content Sets >

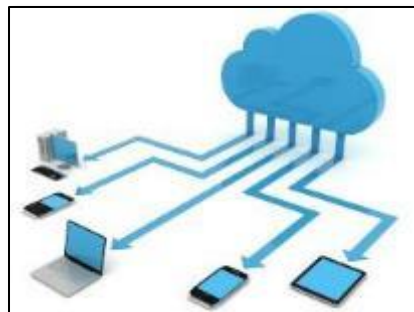
現在 Lab に搭載可能なコンテンツセット

- 17th & 18th Century Burney Collection
- 17th & 18th Century Nichols Collection
- 19th Century UK Periodicals
- American Amateur Newspapers
- American Civil Liberties Union Papers, 1912-1990
- American Historical Periodicals from the AAS
- American Fiction
- Archives Unbound
- Archives of Sexuality & Gender
- Associated Press Collections Online
- Brazilian and Portuguese History and Culture
- British Library Newspapers
- China and the Modern World
- Crime, Punishment, and Popular Culture 1790-1920
- Daily Mail Historical Archive, 1896-2004
- The Economist Historical Archive, 1843-2014
- The Illustrated London News Historical Archive 1842-2003
- The Independent Digital Archive
- Eighteenth Century Collections Online
- Nineteenth Century Collections Online
- Indigenous Peoples of North America
- International Herald Tribune Historical Archive 1887-2013
- Liberty Magazine Historical Archive 1924-1950
- The Listener Historical Archive 1929-1991
- Making of the Modern Law: Foreign Primary Sources
- Making of the Modern Law: Foreign, Comparative and International Law
- Making of the Modern Law: Legal Treatises
- Making of the Modern Law: Primary Sources
- Making of the Modern Law: Trials
- Making of the Modern World
- Nineteenth Century U.S. Newspapers
- Picture Post Historical Archive
- Political Extremism & Radicalism
- Punch Historical Archive, 1841-1992
- Sabin Americana, 1500-1926
- Smithsonian Collections Online
- The Sunday Times Digital Archive 1822-2016
- The Telegraph Historical Archive 1855-2000
- The Times Digital Archive 1785-2012
- Times Literary Supplement Historical Archive 1902-2013
- U.S. Declassified Documents Online
- U.S. Supreme Court Records and Briefs
- Women's Studies Archive

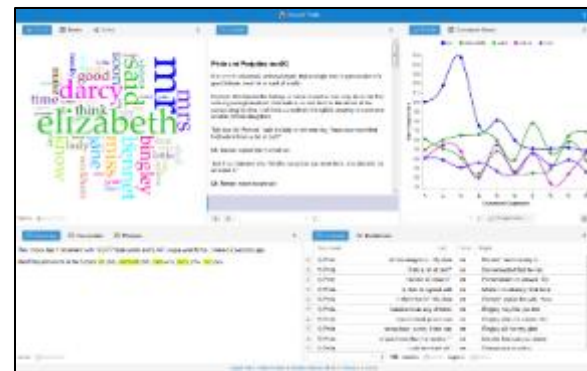
3つの障害を解決



既に分析用に最適化されたデータを格納



クラウド上で管理されている



主な分析ツールを搭載

Gale Digital Scholar Lab デモンストレーション

いよいよ今年！



TOKYO 2020



TOKYO 2020

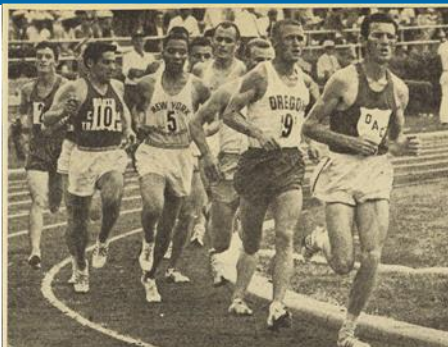
PARALYMPIC GAMES



1964年 東京五輪の海外報道記事をテキストマイニング



TOKYO 1964



5,000-METER ROAD TO TOKYO—U.S. Olympic hopefuls are bunched up during 5,000-meter race in trials in New York. Bob Schul is leading en route to a 14:10.8. Bill Dellinger (9), who finished 2d, Gerry Lindgren (2) was 3d; Oscar Moore (3) 4th. Jim Beatty (10) dropped out with a foot injury and plans an appeal for a World Free International.

JAPANESE SWIMMERS RECEIVE APOLOGY

CRITICISM OF BAN

FROM OUR CORRESPONDENT

DURBAN, FEB. 1

An international "incident" created by the refusal of the Pretoria municipal authorities to allow Japanese swimmers now visiting southern Africa to use a public swimming bath has prompted Mr. Neville Gracie, of Durban, president of the South African Amateur Swimming Union, to apologise to the Japanese swimming management and the Japanese Consul-General in South Africa.

Mr. Gracie today described the ban as "unprecedented in South African sporting



TOKYO. OLYMPIC MEDALS ON DISPLAY.

SPORTING NEWS

NORTH KOREA TO COMPETE IN TOKYO AFTER ALL

PARRY AND AISHER RAISE BRITISH YACHTING HOPES

Tokyo Sept. 30.—The 3,000 or so competitors and officials already in residence at the Olympic villages here were joined by hundreds more today as the huge airlift into Japan continues. The third and last big party of Britain's 270-strong team arrived this morning—members of the boxing, fencing, hockey, shooting, gymnastics and equestrian team.

Overriding all else, however, is the impending row over the banned swimmers and athletes from Indonesia which loomed a little larger today. Competitors in the Games of the New Emerging Forces in Jakarta last year were banned by their various international federations

including the brilliant woman athlete Sin Kim Dan, who took part in G.A.N.E.F.O. Early this month North Korea had told the organizing committee they would not take part in the Olympics because of the ban on G.A.N.E.F.O. athletes and swimmers. Mr. Shigeru Yosano, general secretary of the organizing committee, said he was happy about the North Korean decision to compete. But he added that the organizing committee could not allow banned athletes or swimmers to enter.

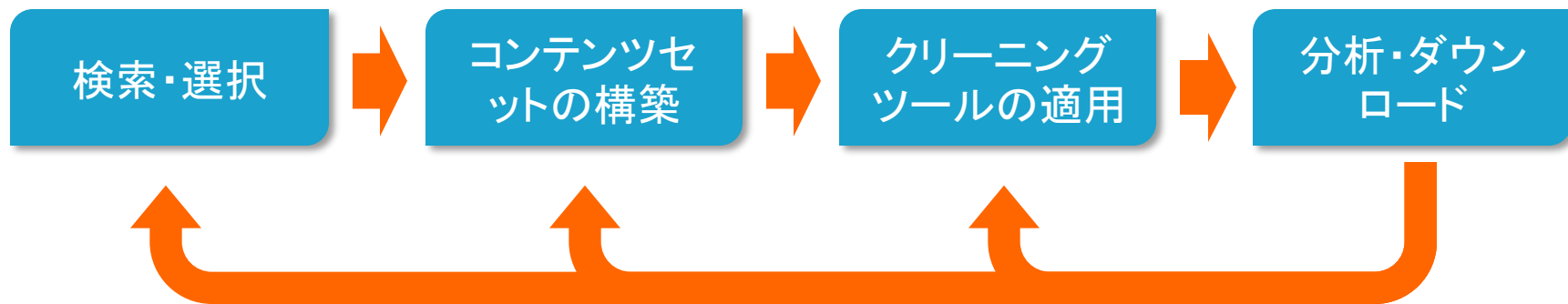
ORDERED TO BED

Mr. R. Hinks, of the British headquarters staff, who was in charge of the party of 85 British competitors who arrived today, said they had an excellent flight. Most



The British coach, A. D. Kinnear, addressing members of Britain's Olympic swimming team in Durban during a break in their first period of training since their arrival.

Gale Digital Scholar Lab のワークフロー



1. 資料を検索する



Advanced

Search for

tokyo olympics

And

And

in

Keyword

ADD A ROW

SEARCH

More Options

by archive

Select

by content type:

by publication

All

Include

Newspapers & Periodicals

1961

and

1961

and

1964

2. 分析対象としたいものを選び集める

DIGITAL SCHOLAR LAB
From Gale

Search using keywords and boolean operator | Basic Search

Search Clean Analyze My Content Sets

Add to Content Set Remove from Content Set Search History

ACTIVE CONTENT SET
Tokyo Olympics 1964
MANAGE | ANALYZE

Search within results

Limit Search by:

Archives

- The Telegraph Histor... (108)
- The Times Digital Ar... (74)
- International Herald... (70)

Content Type

- Newspaper (291)

Document Type

- Article (191)
- Sports news (91)

All Content (291)

Search Terms: Document Title (olympic Or olympics) And Entire Document (tokyo) LIMITS:
Content Type ("Newspaper") And Publication Date (1961 - 1964)

Sort by: Relevance

Select All

1 2 3 NEXT>

Tokyo Pronounced Ready for Olympics ✓ Added
OCR CONFIDENCE 64%
Tokyo Pronounced Ready for Olympics Tokyo Pronounced Ready for Olympics TOKYO, Sept. 10 (UPI).— Daigoro Yasukawa, president of the Tokyo Olympic Games organizing committee, today announced completion of all facilities...

Tokyo Prepares For Olympic Games ✓ Added
BY OUR OWN
OCR CONFIDENCE 92%
TOKYO PREPARES FOR I OLYMPIC GAMES TOKYO PREPARES FOR I OLYMPIC GAMES FROM OUR. OWN CORRESPONDENT TOKYO, MAY 30 The Japanese Emperor has consented to be patron of

ARCHIVE	International Herald Tribune Historical Archive, 1887-2013
SOURCE LIBRARY	The New York Times Company
CONTENT TYPE	Newspaper
DOCUMENT TYPE	Article
PUBLICATION DATE	September 11, 1964
PUBLICATION	New York Herald Tribune (European Edition).
PUBLISHER	
ARCHIVE	The Times Digital Archive
SOURCE LIBRARY	Times Newspapers Limited
CONTENT TYPE	Newspaper
DOCUMENT TYPE	Article
PUBLICATION DATE	May 31, 1962
PUBLICATION	The Times

2. 分析対象としたいものを選び集める



DIGITAL SCHOLAR LAB

From Gale

Search Clean Analyze My Content Sets



Doc Explorer



Remove from Content Set



Cite



Send to OneDrive™



Download



Email



Print

Olympic Games As Timely Spur For Tokyo Development

BY From Our Own Correspondent

Show Source Citation

ACTIVE CONTENT SET

Tokyo Olympics 1964

MANAGE | ANALYZE

DOC EXPLORER

VIEWS

- Original Document Scan
- OCR Text

SEARCH WITHIN

Search within document



OLYMPIC GAMES AS TIMELY SPUR FOR TOKYO DEVELOPMENT

POPULATION SOARS AHEAD OF AMENITIES

From Our Own Correspondent

ALD FISHER
...writes:—
...to find a better
...that abstruse science
...remote from reality, ca
...important practical pro
...ceded by the work o
...Kotbamsted during h
...in anxious to have th
...ating to our unique fi
...ined by modern stati
...tract information whic
...obtained but which o
...had missed. Neithe
...ridge could then suppl
...stian able and willin
...work, but I heard
...him ready to do so. H
...warm opinion of hir
...had struck to the roa
...side a first-class mathe
...would not"—which su
...as the man we wante
...peared.

IL INVESTIGATION
...some theoretical stud
...development of suitab
...with agricultural exp
...in a monumental inve
...heretical foundations
...es, which was publis
...city and which acqui
...title fame. He used h
...studying the field dat
...divers various researc
...laboratories, includ
...numbers of micro
...the consisting of abli

SHIFTING CITY SCENE

What Tokyo lacks in distinction and civic amenity, however, it makes up in its extraordinary vitality, and this gives



Document Text

OCR Confidence: 91% Learn how this text was created

OLYMPIC GAMES AS TIMELY SPUR, FOR TOKYO DEVELOPMENT, POPULATION SOARS AHEAD OF AMENITIES, ---. From Our Own Correspondent, TOKYO, The newcomer to Tokyo from the west often gets a shock when he is first driven from ship or aircraft to his hotel. He wonders when the seemingly endless narrow and dusty streets, lined with rickety houses with open-fronted booths garishly embellished with neon lights and signs, will at length give place to the city proper. A modern building here and there heightens the expectation, but it is an expectation that is never realized in full. For Tokyo, though the largest capital city in the world, with more than 10 million inhabitants, remains basically the "collection of villages" it was once termed. Now it sprawls endlessly around urban nodes of offices, department stores, cinemas, and railway stations; a teeming monster of metropolitan growth, the despair of town planners and civil administrators. SHIFTING CITY SCENE, What Tokyo lacks in distinction and civic amenity, however, it makes up in its extraordinary vitality, and this gives hope that such blemishes are as temporary as the city scene. The average height of Tokyo houses is less than two storeys, and their construction from wood and plaster-the work of carpenters rather than builders-is as flimsy as a cricket pavilion and as easily erected or removed. This fact underlies the remarkable way in which the profile of the busier streets seems constantly to be changing, with old houses disappearing and new ones popping up almost overnight. A symbol of the changing times is the motorized cement mixer, often driven in a manner more appropriate to a sports car than a heavy lorry, which charges about the city between the many modern building sites. Here great steel cages,

文書の
画像

OCRテキスト

3. 集めたものに名前をつける



🔖 Tokyo Olympics 1964

A collection of newspaper articles from 1961-1964 about the Tokyo Olympics (sports section excluded)

Overview

Documents (191)

Analyses

Search History

ARCHIVES USED (9)

International Herald Tribune Historical Archive 1887-2013 (70)
The Telegraph Historical Archive (56)
The Times Digital Archive (29)
The Sunday Times Digital Archive (28)
The Illustrated London News Historical Archive, 1842-2003 (3)

[View More](#)

DOCUMENT TYPE (1)

Article (191)

AUTHORS (50)

By the Associated Press (11)
By United Press International (8)
Daily Telegraph Reporter (8)
OUR OWN (6)
FROM OUR EQUESTRIAN CORRESPONDENT (5)

[View More](#)

SOURCE LIBRARIES (8)

The New York Times Company (70)
Times Newspapers Limited (57)
Telegraph Media Group (56)
Illustrated London News Ltd (3)
The Economist Newspaper Limited (2)

CONTENT TYPE (1)

Newspaper (191)

PUBLICATION TITLE (10)

New York Herald Tribune (European Edition). (70)
The Daily Telegraph. (51)
The Times. (29)
The Sunday Times. (28)
The Sunday Telegraph. (5)

4. クリーニングツールを設定する(任意)



Save As



Test Configuration

CLEAN

Cleaning Configurations are used for

1. Configuring analysis tools to improve results in [Analyze](#)
2. Downloading content sets in [My Content Sets](#)

[+ NEW CONFIGURATION](#)

YOUR CONFIGURATIONS

[Default Cleaning Configuration](#)

[American Fiction](#)

[Fiction FH](#)

[Default Cleaning Configuration copy](#)

[Japanese stopwords lower but no nos or symbols](#)

[Japan in British Editorials](#)

[All boxes ticked HK](#)

[All boxes ticked USDD okinawa](#)

CLEANING CONFIGURATION

Default Cleaning Configuration

This is Gale's default cleaning configuration. Changes to this configuration can only be saved as a new configuration.

TEXT CORRECTION

- All lower case

CHARACTERS

- Remove all extended ASCII characters
- Remove all number characters

SPECIAL CHARACTERS

- Remove all special characters

[Set specific special characters to remove](#)

PUNCTUATION

- Remove all punctuation

STOP WORDS

[LEARN MORE](#)

- Ignore stop words case

[Choose a Starter List](#) [Clear All](#)

a
about
above
across
after
afterwards
again
against
all
almost

5. 分析ツールを適用する



DIGITAL SCHOLAR LAB

From Gale

Search Clean Analyze My Content Sets



Add Tool



Remove



Refresh



Run Selected

ANALYZE

CONTENT SET

Tokyo Olympics 1964

Analysis Tools (5)

Select All

Clustering

TOOL SETUP

5 cluster (Tue Jan 14 06:59:36 EST 2020)

[View](#) | [New tool setup](#)

RUN STATUS

✓ COMPLETED

RESULTS



Scatter Plot

Named Entity Recognition

TOOL SETUP

NER (Tue Jan 14 07:00:26 EST 2020)

[View](#) | [New tool setup](#)

RUN STATUS

✓ COMPLETED

RESULTS



Entities Found

Ngram

TOOL SETUP

RUN STATUS

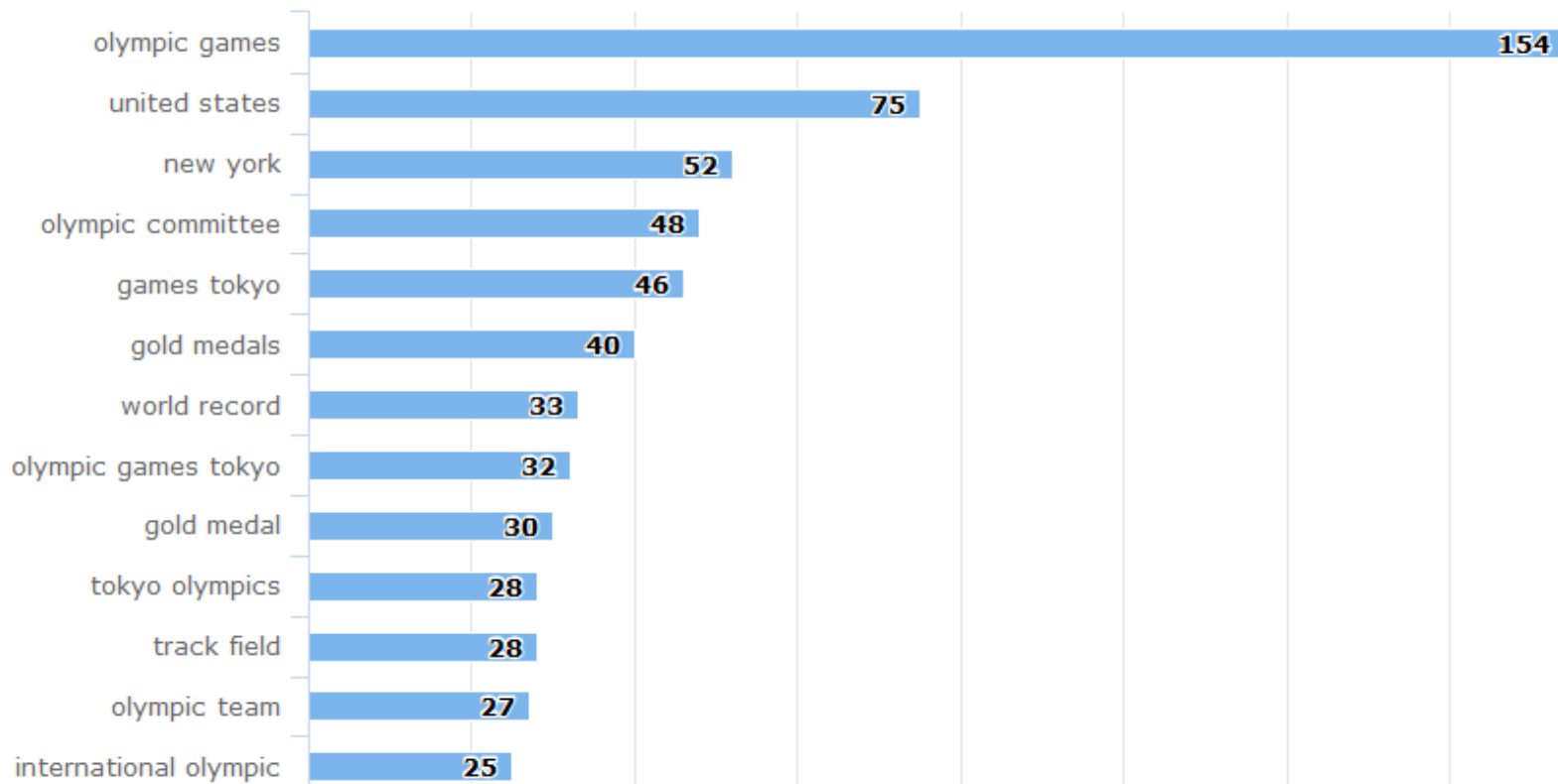
RESULTS

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

ツール名	主な用途	ライブラリ・ツールキット	オープンソース
Nグラム頻度 Ngram	頻出単語・フレーズの抽出	Lucene	×
クラスタリング Clustering	文書を類縁性で分類	SciKit Learn	○
トピック・モデリング Topic Modelling	複数の文書に共通するトピックの抽出	Mallet	○
感情分析 Sentiment Analysis	文書内容が肯定的か否定的かを分析	OpenNLP	○
固有表現抽出 Named Entity Recognition	固有名詞・数字・日付等の抽出	spaCy	○
品詞タグ付け Parts-of-Speech Tagger	著者ごとに品詞の使用頻度を比較	spaCy	○

分析の例：Nグラム頻度（頻出単語・フレーズの抽出）

Term Frequency



分析の例: クラスタリング (文書を類縁性で分類)

< Clustering
5 cluster (Tue Jan 14 06:59:36 EST 2020) ▾ Results
Cluster

⚙️ Tool Setup
📄 Download
ℹ️ About

VISUALIZATIONS

Cluster

● Scatter Plot



分析の例: 固有表現抽出 (固有名詞・数字・日付等を抽出)

ENTITY SEARCH

Search all entities found

RESET

ENTITY CATEGORIES

- DATE
- TIME
- GEOGRAPHY
- GEO-POLITICAL ENTITY
- PLACE
- ARTWORK
- EVENT
- LAW
- PRODUCT
- PERSON
- MEASUREMENT
- MONEY
- NUMBER

Filter entities

Top 200 Entities

Category


Documents

Count

Tokyo	GEO-POLITICAL ENTITY	163	302
first	POSITION	68	150
two	NUMBER	59	114
Japanese	CULTURAL GROUP	46	111
U.S.	GEO-POLITICAL ENTITY	31	97
Olympics	EVENT	55	88
Japan	GEO-POLITICAL ENTITY	44	81
1	NUMBER	29	78
2	NUMBER	19	76
Olympic	GEOGRAPHY	48	75
the Olympic Games	EVENT	57	74
three	NUMBER	40	74

分析の例: 固有表現抽出 (固有名詞・数字・日付等を抽出)

Indonesia Leaves Olympic Body

Document Text 

year DATE , which the I.O.C. ORGANIZATION president, Mr. Avery Brundage PERSON , announced last weekend DATE was because of the " scandalous occurrence at last year's DATE fourth POSITION Asian Games EVENT in Jakarta GEO-POLITICAL ENTITY ". The Indonesian CULTURAL GROUP national Olympic GEO-POLITICAL ENTITY committee announced today DATE it had sent a cable to the I. O.C. GEO-POLITICAL ENTITY renouncing its membership.. The message said the I. O.C. GEO-POLITICAL ENTITY 's suspension of Indonesia GEO-POLITICAL ENTITY " is a violation of the Olympic GEOGRAPHY charter since not a single article in the charter warranted such drastic sanction ". it said "The I. O.C. GEO-POLITICAL ENTITY executive committee has sacrificed the interests of 2,000 million NUMBER people in Asia GEOGRAPHY in order to serve the political manoeuvres of Israel GEO-POLITICAL ENTITY and Formosa". Mr. Brundage had said the I.O.C. ORGANIZATION decision was final and could only be rescinded if Indonesia guaranteed no political discrimination in sports in future. Indonesia GEO-POLITICAL ENTITY barred Formosa GEO-POLITICAL ENTITY and -Israel, with whom sbe has no diplomatic relations, from the Asian Games EVENT in Jakarta GEO-POLITICAL ENTITY last August DATE NEW GAMES President ORGANIZATION Sukarno PERSON last night TIME ordered the Sports Minister. Mr. Maladi PERSON , to arrange Indonesia GEO-POLITICAL ENTITY 's withdrawal from the I.O.C. GEO-POLITICAL ENTITY and at the, same time to organize " Games of the new emerging forces" (Ganefo). Dr. ARTWORK Sukarno PERSON said at the opening session of the All Indonesia GEOGRAPHY . National Front ORGANIZATION conference here that Indonesia GEO-POLITICAL ENTITY would invite the newly independent and unaligned nations in Asia GEOGRAPHY , Africa GEOGRAPHY , and Latin America GEOGRAPHY to take part in the proposed games with countries belonging to the socialist block He said sport and politics were " inseparable " pointing out that communist China PERSON was excluded from the Asian Games

Named Entity Recognition

- > DATE (6)
- > TIME (1)
- > GEOGRAPHY (5)
- ✓ GEO-POLITICAL ENTITY (8)
 - ✓ Indonesia (6)
 - ... suspension of Indonesia from the OL...
 - ... suspension of Indonesia " is a violatio...
 - ... sports in future.Indonesia barred Fo...
 - ... Maladi, to arrange Indonesia's withdr...
 - ... here that Indonesia would invite the...
 - ... also ordered Indonesia's withdrawal ...
- > I.O.C. (1)
- > Jakarta (2)
- > Formosa (1)
- > Israel (1)
- > O.C. (3)
- > Olympic (1)
- > Tokyo (1)
- > ARTWORK (2)
- > EVENT (3)

分析の例:感情分析 (内容が肯定的か否定的かを分析)

docid	document title	publication year	sentiment score	relative mean score
LYVXAB289829111	U. S. Jumpers May Surprise in Olympics	1962	2.5	2.5372
FP1803030825	Ann Packer, Olympic bride with a golden rose	1964	2	1.9063
HYUCFP511490363	Cawley Breaks World Mark at Olympic Qualifications	1964	1.9444	1.8362
CS202729299	Olympic Excitement In Welsh Village	1964	1.9231	1.8092
FP1802662281	Andromeda is chosen for Tokyo Olympics	1964	1.9062	1.788
FP1800544728	Wiggs seems certain of an Olympic place	1964	1.8333	1.696
IO0706778356	Olympic ?gold? for Mary Rand	1964	1.8125	1.6697
CS118843205	Olympic Riders At Wembley	1964	1.7857	1.6359
UPBWPI500799500	Tokyo Olympics Finished; Called History's Best Games	1964	1.75	1.5909
FP1800686690	Olympic preparations	1963	1.75	1.5909
IO0706779740	Pupils cheer their teacher's Olympic win	1964	1.7273	1.5622
CS252667689	Olympic Pointer At Burghley	1964	1.6667	1.4857
FP1802518077	Hint of Olympic gold from canoeists	1964	1.5769	1.3725
FP1800541554	British Olympics team face stiff test	1964	1.5667	1.3596
MCIDWE996480541	Strong Showing of Weak U.S. Team in Sao Paulo Boosts Olympic Hopes	1963	1.5581	1.3488
IO0704234472	Government Gives ¥10,000 for Olympics	1964	1.5	1.2754
EPSNGN926349544	U.S. Olympic Picks Gather	1964	1.5	1.2754
IO0704227741	Sport named on Olympic medals	1964	1.5	1.2754
YSOIAH104602501	Kennedy Backs Detroit Efforts for '68 Olympics		1.5	1.2754
FP1800694195	Sprinters win an Olympic repeal	1964	1.5	1.2754
FP1800697202	Olympic party's stiff programme	1963	1.4444	1.2052

分析の例:トピック・モデリング (共通するトピックの抽出)

Logout My Library Links English Signed in: Masaki

VIEWS

- Topic overview
- Topic Comparison

Topic 0

IDENTIFIED IN

36 DOCUMENTS

TOPIC MEASURES

Tokens	2124
Document Entropy	3.5193
Average Word Length	6
Coherence	-42.9297
Uniform Distance	2.7596
Corpus Distance	2.3072
Exclusivity	0.7472

Topic 1

IDENTIFIED IN

29 DOCUMENTS

TOPIC MEASURES

Tokens	2686
Document Entropy	3.0264
Average Word Length	4.7
Coherence	-41.7343
Uniform Distance	2.5033
Corpus Distance	1.9715
Exclusivity	0.6483

Topic 2

IDENTIFIED IN

TERMS	COUNT	PROBABILITY	DOCS
television	45	0.0212	18
satellite	40	0.0188	17
pictures	35	0.0165	16
syncom	27	0.0127	10
states	26	0.0122	16
bbc	22	0.0104	7
united	21	0.0099	13
use	14	0.0066	9
day	14	0.0066	11
europe	14	0.0066	11

TERMS	COUNT	PROBABILITY	DOCS
team	33	0.0123	15
won	29	0.0108	12
horse	27	0.0101	13
second	27	0.0101	12
riders	24	0.0089	11
horses	22	0.0082	10
david	19	0.0071	8
miss	19	0.0071	8
cup	18	0.0067	9
event	16	0.006	9

TERMS	COUNT	PROBABILITY	DOCS
japanese	42	0.0151	18

分析の例:トピック・モデリング (共通するトピックの抽出)

Logout My Library Links English Signed in: Masaki

TOPICS

- Broadcasting
- Horseracing
- Ceremony
- Measurements
- Athleticism
- Politics
- General
- Records
- Random
- Japan

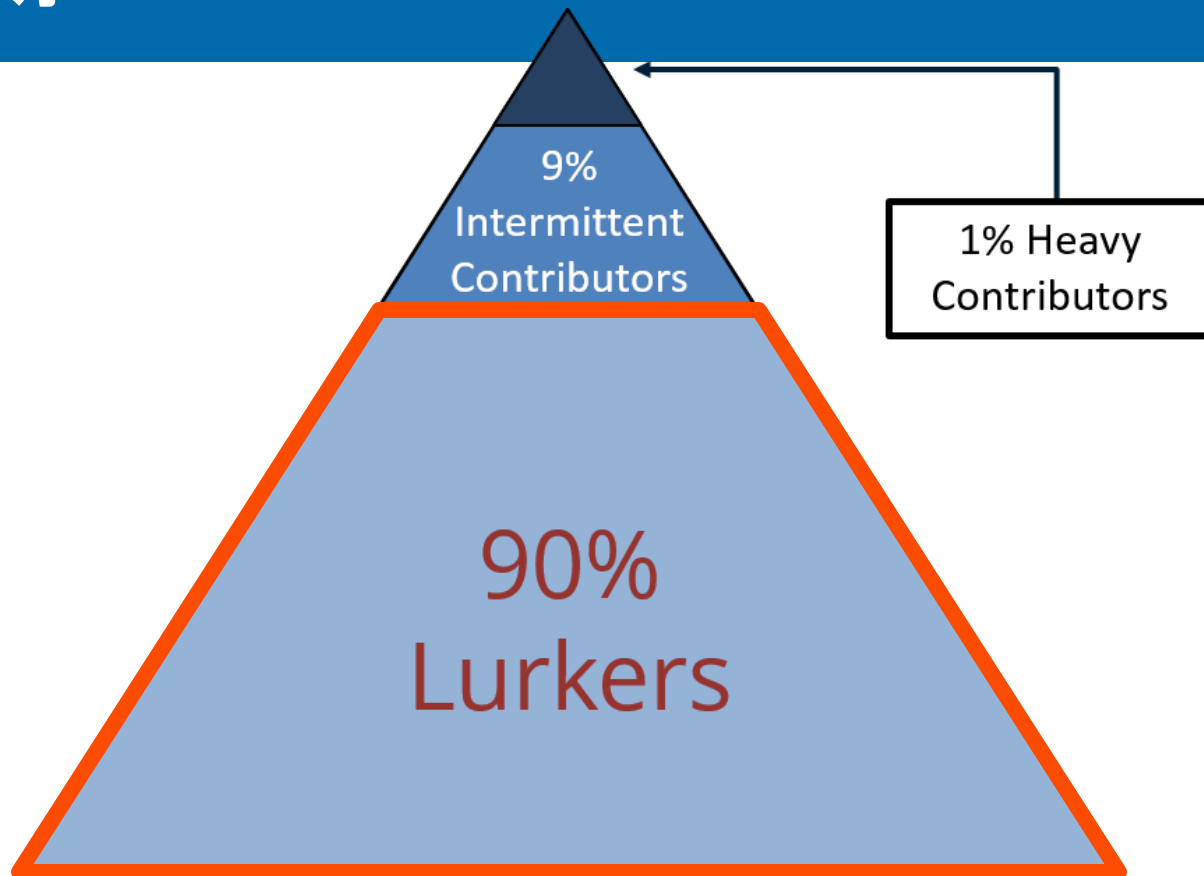
DOCUMENTS DISPLAYED (14)

Select Documents

Topic Proportion



‘1-9-90’ の法則



The 90-9-1 rule for participation in
an online community
[http://www.nngroup.com/articles/p
articipation-inequality/](http://www.nngroup.com/articles/participation-inequality/)

実際の応用例



W
UNIVERSITY of
WASHINGTON

A screenshot of a digital humanities project page. The page has a dark header with navigation links and a search bar. The main content area features a large, stylized graphic with the title "Now CHOP SUEY at HOME" and various text snippets from historical newspaper articles. To the right of the graphic is a short paragraph of text. The overall design is clean and modern, typical of a digital humanities interface.

in: Student Exhibits, Evolving Attitudes Toward Chinese Food in the U.S., The Terrace Room in WWII, By the Glass, Vegetarianism in WWII

Learning across the years: Food Patterns in Boom and Bust, The Journey of Eggs Benedict

Intro to Digital Humanities Fall 2018

Sections Timeline

- Chop Restaurant Syndrome
- Chop Suey
- Eat Chop Food Products
- Women's Visit to China
- Class Data Visualizations
- About
- Credits

Evolving Attitudes Toward Chinese Food in the U.S.

Now CHOP SUEY at HOME

Join us guys go off in Japan
DENSE RESTAURANTS
Chinese food products
your crazy MSG
is NO, I respect
the necessary course
and electricity. American states
the suspicious American press
The are fully enclosed by laws
the Douglas and Good House
The Young Business of Food.

As an analysis of how Chinese food was viewed in the United States over time through exploration of documents and photos that cover the Chinese restaurant landscape as an extension of anti-Chinese sentiment, Chop Suey and Lo Chop food products as anti-Chinese food was associated, and finally Robert's 1972 visit to China and its impact on the perception of Chinese food in America.

ワシントン大学「デジタル人文学入門」

2015～2018年に開講した共通科目

- 初めての開講、学部生・院生の両方が履修可能
- 週2回、各回2時間、単学期(10週)
- 受講生の学年、所属学科はさまざま



ワシントン大学 サラ・ケッチリー先生

Sara L. Ketchley, "Introducing Digital Humanities in the Undergraduate Classroom: Strategies, Solutions, and Pedagogical Practices Using the Gale Digital Scholar Lab" in *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching*, Farmington Hills, MI: Gale, 2020

ワシントン大学「デジタル人文学入門」

様々な技術的困難：

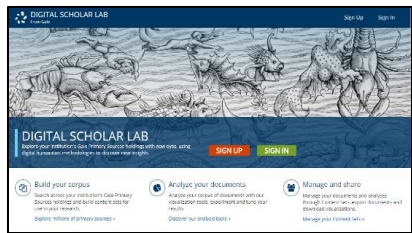
- ソフトウェアのインストールとトラブル対応で貴重な授業時間が費やされてしまう
- さまざまなプラットフォームがあり、それぞれに癖があるので、使い慣れるまでに時間がかかる
- 大学側のITサポートも期待できない



ワシントン大学 サラ・ケッチリー先生

Sara L. Ketchley, "Introducing Digital Humanities in the Undergraduate Classroom: Strategies, Solutions, and Pedagogical Practices Using the Gale Digital Scholar Lab" in *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching*, Farmington Hills, MI: Gale, 2020

ワシントン大学「デジタル人文学入門」



ワシントン大学 サラ・ケッチリー先生

Gale Digital Scholar Lab の導入によって:

- コーパスの作成、データのキュレーション、分析、視覚化までを一つのプラットフォーム上で行えるようになった。
- クラウドベースなので、端末の環境に左右されなかった
- 研究課題の掘り下げや結果のプレゼンテーションにより注力できるようになった。

Sara L. Ketchley, "Introducing Digital Humanities in the Undergraduate Classroom: Strategies, Solutions, and Pedagogical Practices Using the Gale Digital Scholar Lab" in *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching*, Farmington Hills, MI: Gale, 2020

学生による発表の例:

「アメリカ人の中華料理に対する態度の変遷」



ワシントン大学 サラ・ケッチリー先生

The screenshot shows a digital humanities exhibit page. At the top, there is a navigation bar with links: "Browse Student Exhibits", "Evolving Attitudes Toward Chinese Food in the U.S.", "The Terrace Room in WWII", "By the Glass", and "Vegetarianism in WWI". Below this is a search bar and the text "Intro to Digital Humanities Fall 2018". The main content area is titled "Evolving Attitudes Toward Chinese Food in the U.S.". On the left, there is a sidebar menu with the following items: "Interactive Timeline", "Chinese Restaurant Syndrome", "Chop Suey", "La Choy Food Products", "Nixon's Visit to China", "Class Data Visualizations", "About", and "Credits". The main content area features a large, stylized collage of historical newspaper clippings. The most prominent headline reads "NOW CHOP SUEY AT HOME". Other visible text includes "Just as you get it in famous CHINESE RESTAURANTS", "Chinese food make you crazy? MSG is No. 1 suspect", and "Choy ingredients are in sanitary conditions". Below the collage is a photograph of a man in a suit eating at a table. To the right of the collage, there is a text block: "An analysis of how Chinese food was viewed in the United States over time through exploration of documents and articles that cover the Chinese restaurant syndrome as an extension of anti-Chinese sentiment; Chop Suey and La Choy Food Products as ways Chinese food was Americanized; and finally Nixon's 1972 visit to China and its impact on the perception of Chinese food in America."

<http://www.newbookdigitaltexts.org/dh-fall2018/exhibits/show/chinese-cuisine>

講義を振り返って

- 学生が批判的思考、分析的推論、効果的なコミュニケーションなどの能力を身につけたのはもちろん、
- データとの格闘や管理、プロジェクトの立案と運営、多様な媒体のための文章執筆、データ分析と視覚化、チーム研究とコラボレーションなどの具体的なスキルを身につけた
- → 卒業後も就職や実社会で役に立つと学生自身が実感



ワシントン大学 サラ・ケッチリー先生

Sara L. Ketchley, "Introducing Digital Humanities in the Undergraduate Classroom: Strategies, Solutions, and Pedagogical Practices Using the Gale Digital Scholar Lab" in *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching*, Farmington Hills, MI: Gale, 2020

まとめ

- 一次資料のデジタル化が可能にする新たな研究
 - 〈検索⇒閲読〉 〈収集⇒分析〉の2つの異なるワークフロー
- デジタル人文学実践の3つの障害
 - 分析に適したデータの入手、その保存・管理、分析ツールの難しさ
- 一部の限られた研究者から、より幅広い層へ広がる可能性
 - デジタル時代の人文学教育に大きな示唆

ご清聴ありがとうございました