

Building Data Collections for Digital Humanities Research

Devin Higgins
MSU Libraries
higgi135@msu.edu
[@devinhhi](#)



"humanities data"

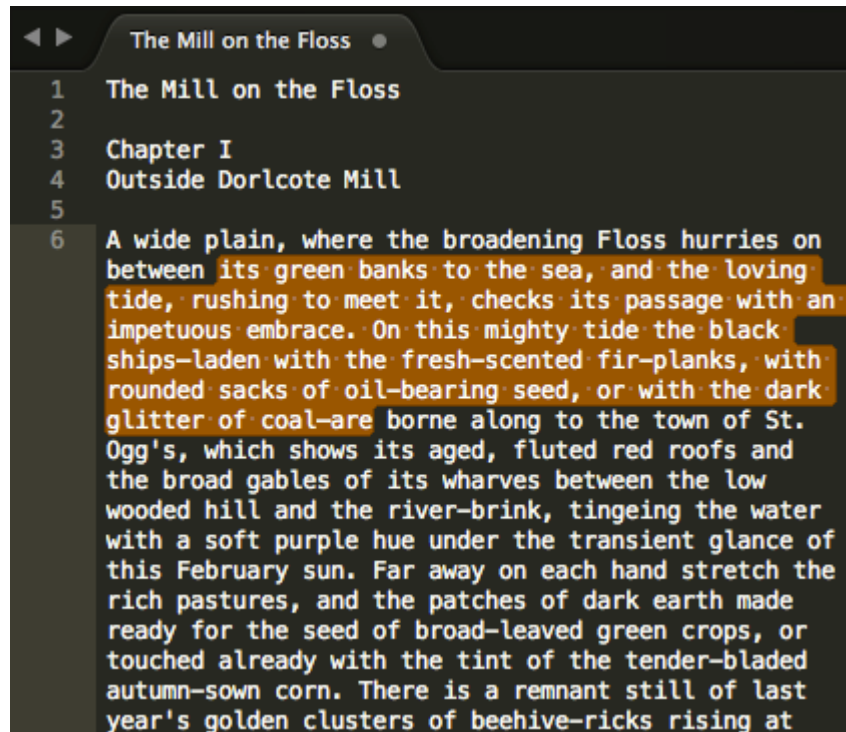
Google Ngrams: "Humanities Data"

THE MILL ON THE FLOSS

CHAPTER I

OUTSIDE DORLCOTE MILL

A WIDE plain, where the broadening Floss hurries on between its green banks to the sea, and the loving tide, rushing to meet it, checks its passage with an impetuous embrace. On this mighty tide the black ships—laden with the fresh-scented fir-planks, with rounded sacks of oil-bearing seed, or with the dark glitter of coal—are borne along to the town of St. Ogg's, which shows its aged, fluted red roofs and the broad gables of its wharves between the low wooded hill and the river brink, tinging the water with a soft purple hue under the transient glance of this February sun. Far away on each hand stretch the rich pastures, and the patches of dark earth, made ready for the seed of broad-leaved green crops, or touched already with the tint of the



From readable page to computable text

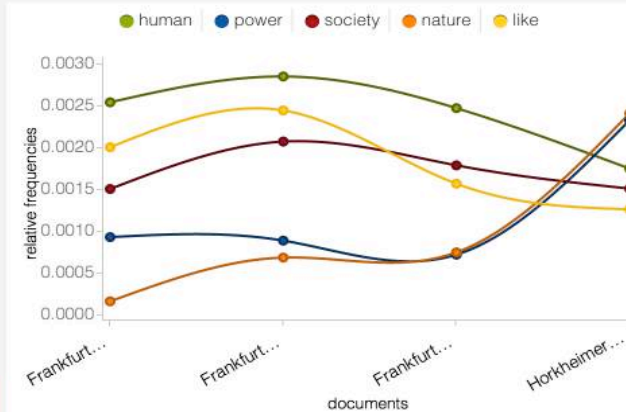


Frankfurt School: Minima Moralia by Theodor Adorno 1951

Theodor Adorno

Minima Moralia Dedication

The melancholy science, from which I make this offering to my friend, relates to a realm which has counted, since time immemorial, as the authentic one of philosophy, but which has, since its transformation into method, fallen prey to intellectual disrespect, sententious caprice and in the end forgetfulness: the teaching of the good life. What philosophy once called life, has turned into the sphere of the private and then merely of consumption, which is dragged along as an addendum of the material production-process, without autonomy and without its own substance.



relative ▾

- This corpus has 4 documents with 213,176 total words and 21,582 unique word forms. Created about 21 minutes ago.
- Documents ordered by number of words (↕):
 - Horkheimer_Max_Adorno_The... (121,119)
 - Frankfurt School: Minima... (36,500)
 - Frankfurt School: Minima... (29,528)
 - Frankfurt School: Minima... (26,029)
- Documents ordered by vocabulary density (↘):
 - Frankfurt School: Minima... (0.206)
 - Frankfurt School: Minima... (0.199)
 - Frankfurt School: Minima... (0.186)
 - Horkheimer_Max_Adorno_The... (0.135)

Document	Left	Term	Right
3) Frankfur...	later redeemed in things and	human	beings. Such anticipation, satiated in
1) Frankfur...	by the social order, could	human	beings introduce a more humane
3) Frankfur...	the road, the breathless flight.	human	dignity insisted on the right
2) Frankfur...	of something unbearable, which befalls	human	beings: debasement. The dream of
3) Frankfur...	of being pursued. Whenever however	human	beings are told: "run," ranging
1) Frankfur...	itself over the heads of	human	beings, indeed by virtue of
2) Frankfur...	stork brings children from. – Every	human	being has an archetype out

451 context
 expand

"non-consumptive research" is text mining or computational "analysis...on one or more Books, but not research in which a researcher reads or displays substantial portions of a Book to understand the intellectual content presented within the Book."

*Google v. the Authors Guild and the Association of American Publishers
Settlement Agreement, 2009*

TECHNOLOGY

After 10 Years, Google Books Is Legal

Thanks to a landmark ruling, information just got a little more free.

4.1k



355



ROBINSON MEYER | OCT 20, 2015

The Atlantic

Google Books Dataset

Browse Topic

English literature	36
Authorship	4
Carpenters	3
English fiction	3
Illegitimate children	3
Infanticide	3
Triangles (Interpersonal relations)	3
Women clergy	3
Christianity	2
English essays	2

Publication Year

1900	75
1907	60
1908	52
1910	40
1884	38
1901	34
1880	29
1886	28
1893	28
1878	26

Search Results

744 Volumes

[NEW SEARCH](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [30](#)

1. Romola ...

[TEXT](#) [METADATA](#) [HATHITRUST](#) [COLLECT](#)

Main Author: [Eliot, George, 1819-1880](#).
ID: [njp.32101068597556](#)

2. The complete works of George Eliot.

[TEXT](#) [METADATA](#) [HATHITRUST](#) [COLLECT](#)

Main Author: [Eliot, George, 1819-1880](#).
ID: [nnc1.0047924500](#)

3. Romola ...

[TEXT](#) [METADATA](#) [HATHITRUST](#) [COLLECT](#)

Main Author: [Eliot, George, 1819-1880](#).
ID: [njp.32101068597564](#)

4. Scenes of clerical life, by George Eliot. (1800)

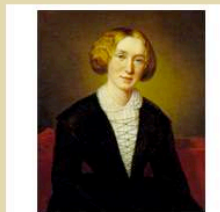
[TEXT](#) [METADATA](#) [HATHITRUST](#) [COLLECT](#)

Main Author: [Eliot, George, 1819-1880](#).
Series: [Novels of George Eliot](#).
ID: [mdp.39015010470675](#)

5. The poems of George Eliot. (1800)

[TEXT](#) [METADATA](#) [HATHITRUST](#) [COLLECT](#)

Main Author: [Eliot, George, 1819-1880](#).
Add'l Author: [Rands, W. B. \(William Brightly\), 1823-1882](#).
ID: [wu.89080367642](#)



George Eliot

[DBpedia](#) [VIAF](#) [Wikipedia](#) [LOC](#)

Download Options

Corpus size currently limited to 500 volumes (= 500MB).

- Download collected volumes
- Create subset of volumes randomly drawn from all results.
- Use all results.

Download:

- Full text.
- Bibliographic metadata (Title, Author, etc.)
- Technical metadata

[Download Text & Metadata](#)

Create list of unique identifiers for all 744 volumes and download.

[Submit](#)



Sunday School Books in Nineteenth Century America

(open to non-MSU users)

Number of Works: 166 works

Years covered: 1809-1887

Size: 11.6 MB



The Grange Visitor

(open to non-MSU users)

Number of Works: 429 issues

Years covered: 1875-1896

Size: 8.53 GB



Feeding America

(open to non-MSU users)

Number of Works: 76 books

Years covered: late 18th - early 20th century

Size: 78+ MB



U.S. Congressional Collection

(open to on-campus users)

Number of Works: 17,000+ daily records

Years covered: 1789-1918

Size: 120+ GB



Google Books Dataset

(open to on-campus users)

Number of Works: 3,000,000 approx.

Years Covered: 1500 - 2012

Size: 2.9 TB

library originated data

purchased data

negotiated data

THE GRANGE VISITOR

ISSUED SEMI-MONTHLY,

EXECUTIVE COMMITTEE



BY ORDER OF THE

MICHIGAN STATE GRANGE.

Vol. 9.—No. 4. }
WHOLE No 84 }

SCHOOLCRAFT, FEBRUARY 15th, 1880.

[Subscription Publishing Co.'s Print.]

{ YOUR SUBSCRIPTION
{ will Expire with No.

Entered at the Post Office at Schoolcraft as Second Class matter.

THE GRANGE VISITOR,

Is Published on the First and Fifteenth of every Month

AT FIFTY CENTS PER ANNUM,
Invariably in Advance.

J. T. COBB, Editor and Manager.

To whom all communications should be addressed, at Schoolcraft, Mich.

Remittances should be by Registered Letter, Money Order or Draft.

TERMS OF ADVERTISING:

Acceptable advertisements inserted at the rate of \$1.00 per square, for each insertion. A Liberal discount will be made on standing advertisements of three months or more.

INDEX TO THIS NUMBER.

The Silver Lining.—The Transportation Question.—Eaton County Farmers' Institute.—1
The Bachelor's Last Hours.—Care and Feeding of Stock.—The Model Granger.—A Grange Boom.—A Christmas Carol. 2
Sustaining Agricultural Fairs.—A Granger's Opinion of U. S. Patent Laws.—A Good Candidate.—The Governor Question.—Comparative Nutritional Matter in Food.—The Hand that Rules the World.—Final Report to the Farmers' Mutual Defense Association.—Collecting Statistics of Farm Productions.—A Was Father.—Guinea Hen.—Horse Life. 3
Battle Creek Co-operative Store, and Climax Institute.—Kalamazoo Potatoes Grange, No. 27.—Copy.—Bound for California.—State Grange Proceedings.—Picking by the Way. 4
Lecturer's Appointments.—Mason Grange and its Work.—Notice of Meetings.—Our Detroit Agency. 5
Rhymes by (Auntie) Anti-Tobacco.—Care of the

THE SILVER LINING.

There's never a day so sunny
But a little cloud appears;
There's never a life so happy
But has had its time of tears;
Yet the sun shines out the brighter
When the stormy tempest clears.

There's never a garden growing
With roses in every plot;
There's never a heart so hardened
But has one tender spot;
We have only to prune the border
To find the forget-me-not.

There's never a cup so pleasant
But has bitter with the sweet;
There's never a path so rugged
That bears not the print of feet;
And we have a halper promised
For the trials we may meet.

There's never a sun that rises
But we know 'twill set at night;
The hints that gleam in the morning,
At evening are just as bright;
And the hour that is the sweetest
Is between the dark and light.

There's never a dream that's happy
But the waking makes us sad;
There's never a dream of sorrow
But the waking makes us glad;
We shall look some day with wonder
At the troubles we have had.

There's never a way so narrow
But the entrance is made straight,
There's always a guide to point us
To the "little wicket gate;"
And the angels will be nearer
To a soul that is faithful.

Master's Department.

J. J. WOODMAN. - - - PAW PAW

The Transportation Question.

It is well known to every Patron of Husbandry in the land that the Na-

No sooner was the memorial presented to Congress than opposition to the movement began to manifest itself among railroad officials and attorneys, who flocked to Washington to look after the interests of their roads and their clients. Merchants, and other business men whose success depends upon the prosperity of agriculture, are beginning to look this question square in the face, and with few exceptions strike hands with the "Grangers," and are enrolling themselves with the thousands that are now petitioning Congress for National intervention and control of inter-State commerce. There are some of the leading political papers of all the different political parties now daily advocating the measure, and Boards of Trade are not idle lookers-on in this most important movement ever inaugurated in the interest of the producers and consumers in this country; but business men as they are, they are looking to their own interests, and not specially to the interests of farmers or other business men whose interests are inseparably connected with agriculture.

The New York Board of Trade and Transportation have appointed a committee to investigate this Grange movement and examine the Grange petitions and bills now pending before Congress, and recommend such legislation as the "Board" may deem to be for the best interests of the

Board of Trade does not think "practicable or advisable," but thinks the Reagan bill "more wise, as it leaves the railroads free to fix any rates they please," as the following extract from the report of their committee will show:

"We do not think it practicable or advisable, with the knowledge at the disposal of the public at this time, to attempt to fix rates on inter-State traffic, although there can be no doubt of the power of Congress to do so. It is attempting too much with our present knowledge, and would probably result in failure. In this respect, the Reagan bill, now pending in Congress is much more wise, as it leaves the railroads free to fix any rates they please, only stipulating that they shall be public, and alike to all, and that no more shall be charged for a short than for a long distance."

Although the farmers of the country, who are now being "robbed of their well-earned profits," by the transportation companies, will not be easily convinced that there should be no limit to the amount which the companies may have a right to charge for transporting their surplus products to market, yet they will rejoice to know that such influential business men as compose the "New York Board of Trade and Transportation," are using their influence to protect the business interests of the country, in some degree, from the impending danger with which they are threatened by the usurping power of the railroad combinations, of which the National Grange has warned

Farmers' Institute—Eaton County.

VERMONTVILLE, Feb. 8, 1880.

Brother J. T. Cobb:

I send you these items from Eaton County for our Visitor:

The adjourned meeting of the Eaton County Quarterly Convention, F. of H., met at Ketchum's Hall, Charlotte, Jan. 14. Members were present from nearly every Orange in the County, although the traveling was very bad. The principal topic of discussion was the formation of a Pomona Grange in this County. After a full interchange of views, it was voted, nearly unanimously to have one organized. A petition was then circulated, and sent to Worthy Master Woodman, who has appointed Feb. 18, at 1 o'clock, P. M. as the time, and we have Ketchum's Hall, Charlotte, as the place of organization. We now hope and trust that Eaton County will wheel into line with her sisters in the work of advancing this good and glorious cause—the union and elevation of farmers.

On the evening of Jan. 14, and on the following day and evening, a Farmers' Institute was held, under the auspices of the County Agricultural Society, several interesting papers were read, and no aid was received from outside the County. Bro. Ezek. Pray, of Windsor Grange, No. 619, presided and Bro. Seth Ketchum, of Charlotte,

MSU LIBRARIES

[General Info](#)
[Resources](#)
[Research Help](#)
[Using the Library](#)
[Computers](#)
[Contact](#)

Digital Humanities

[About](#)
[Events](#)
[Humanities Data](#)
[Sunday School Books](#)
[The Grange Visitor](#)
[Feeding America](#)
[United States Congressional Collection](#)
[Google Books](#)
[Consultation](#)
[Guide](#)

Related Services

[Digital Scholarship Collaborative](#)

Humanities Data

MSU Libraries Humanities data includes but is not limited to digitized and born digital text, audio, images, moving images, and the metadata that describes them. Current collection strengths reside in text and audio data. The following collections have been prepared with an eye toward enabling computational analysis at the micro and macro scale.

Recommendations for acquisition of new Humanities data, requests for assistance gathering and preparing Humanities data, and questions about how to use Humanities data should be directed to the MSU Libraries [Digital Humanities](#) team.

Text



Sunday School Books in Nineteenth Century America

(open to non-MSU users)

Number of Works: 166 works

Years covered: 1809-1887

Size: 11.6 MB



The Grange Visitor

(open to non-MSU users)

Number of Works: 429 issues

Years covered: 1875-1896

Size: 8.53 GB

lib.msu.edu/dh/humdata

Thomson Reuters Web of Science

*"Lite" (free) API to return metadata,
"Premium" (costly) API to return citation data.*

JSTOR

*Web-Accessible word-frequencies and metadata, full
text by request/contract*

Gale

*Data available for purchase on "cost-recuperative"
basis*

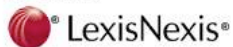
APIs and Bulk Text Access

1	Timestamp	Series	Collection	(text) Collection Size	indicate scale	(image) Collection size	indicate scale	Collection Description
5	9/29/2015 12:54:30		Sabin Americana: 1500-1926 (backfile through 2010)	147 GB		1673 GB		Based on Joseph Sabin's famed bibliography, Bibliotheca Sabiniana, this collection covers a span of 400 years in North, Central and South America. This digital collection consists of testimony and documents.
6	9/29/2015 13:02:22	Archives Unbound	Personal Justice Denied: Public Hearings of the Commission on the Causes of the Current Disfranchisement	51 MB		1.8 GB		The handwritten documents within Personal Justice Denied are part of a larger collection, Part IV: The Age of Emancipation, the fourth installment of the series.
7	10/5/2015 15:50:08	Slavery and Anti-Slavery: Archives Unbound	Part IV: Age of Emancipation	7 GB		1823 GB		In its entirety, Slavery and Anti-Slavery will consist of more than 100 collections. This collection allows you to explore the internal organization of the organization.
8	10/5/2015 16:03:28	Archives Unbound	Federal Response to Radicalism in the 1960s	0.903 GB		50 GB		This collection allows you to explore the internal organization of the organization. Users have access to the complete run of The Times Literary Supplement from 1945 until 1974, writings and influential criticism of hundreds of books.
9	10/5/2015 16:55:30		TLS Historical Archive - 1902-2006	15 GB		344 GB		The value of the TLS Historical Archive lies in its extensive collection of books. This unique digital collection offers thousands of book reviews.
10	10/8/2015 9:53:06	Archives Unbound	Fannie Lou Hamer: Papers of a Civil Rights Activist	0.278 GB		13 GB		Date Range: 1966-1978 Content: 28,577 pages Source Library: Amistad Research Center Description Fannie Lou Hamer was an voting rights activist and civil rights leader. The multi-disciplinary range of content is the largest and most ambitious project of its kind, Slavery and Antislavery.
11	10/8/2015 10:44:55	Slavery and Antislavery	Part II: Slave Trade in the Atlantic World	22 GB		281 GB		Part II, The Slave Trade in the Atlantic World continues the series.
12	10/9/2015 12:40:04	Archives Unbound	The Papers of Amiri Baraka, Poet Laureate of the Black Panther Party	0.19 GB		12 GB		Date Range: 1960-1988 Content: 10,000 pages Source Library: Auburn Avenue Research Library, Atlanta Description Amiri Baraka is the author of over 40 books of essays, poetry, and fiction.

Or use local hard drives...



Powered by



User ID: schanzme

Search

My Corpus Download Queue

Logout

Search

Instructions: Select source(s), construct search, and preview results. If preview results look appropriate, assign a Corpus Name and queue for download. Downloading is queued for nights and weekends. Please plan ahead to ensure sufficient time to retrieve results. For more guidance consult the [step by step instructions](#).

Available Sources:

No Publication Type(s) Selected ▾

Apply Publication Type Filter

No Title(s) Selected ▾

Add Titles

Selected Title Ids:

Search Method:

Boolean

Date Range:

From: 10/22/2015

To: 10/22/2015

No Date Range

Search Query:

Preview Results

Thanks!

Building Data Collections for Digital Humanities Research

Devin Higgins
MSU Libraries
higgi135@msu.edu
[@devinhhi](#)