

Reading

What Middletown Read

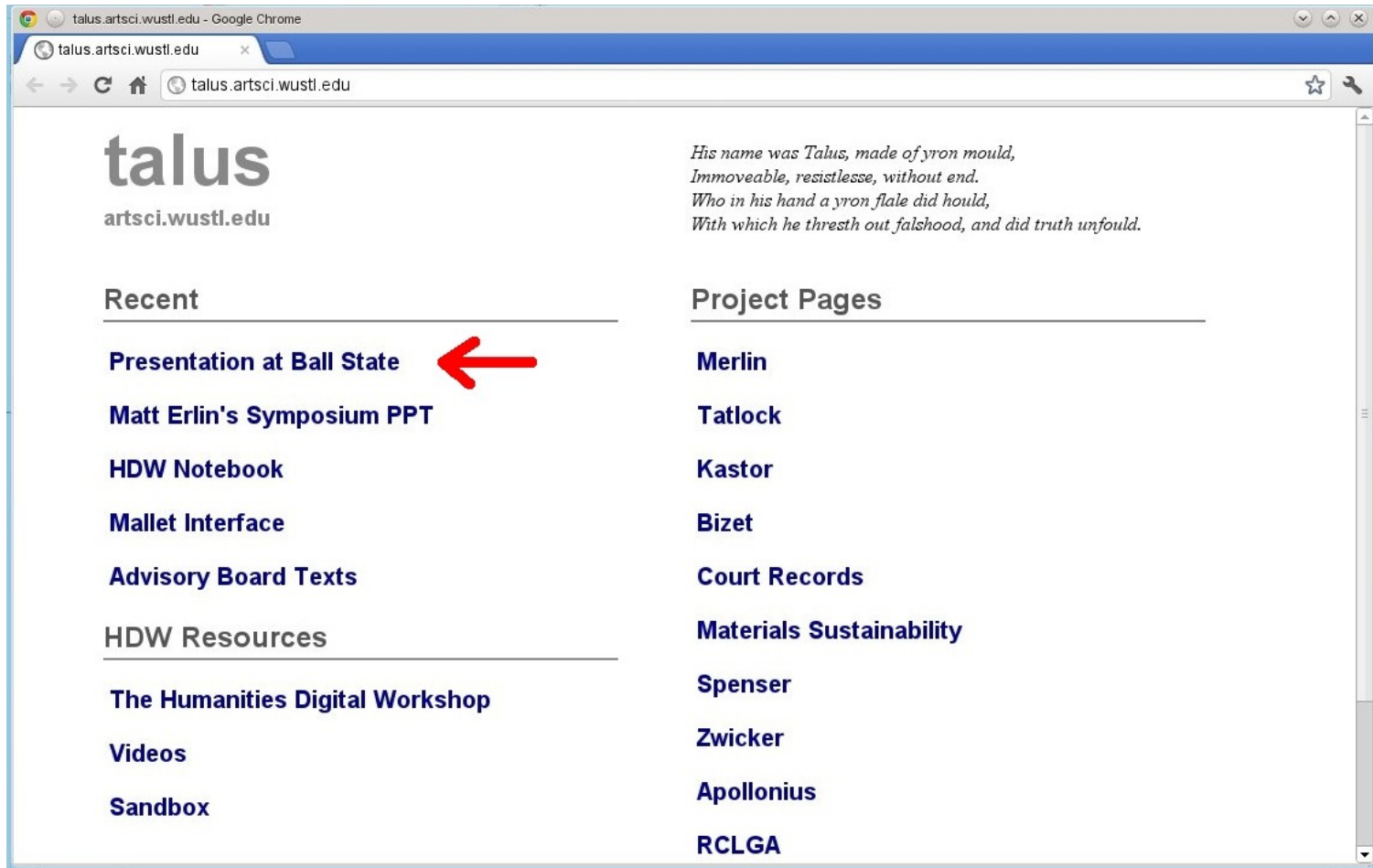
Stephen Pentecost
Humanities Digital Workshop
Washington University in St. Louis
April 18, 2012

The wise library manager, like the children of this world, will hold out as many seductions as possible. Encourage dalliance by scattering about temptations. Pendleton, A.M. “How to Start Libraries in Small Towns—IV.” *The American Library Journal* I, 1877.



Washington University in St. Louis

http://talus.artsci.wustl.edu/



Who are the Humanities Digital Workshop?

Joseph Loewenstein

Professor of English

Director, Interdisciplinary Project in the Humanities

Co-director, Humanities Digital Workshop

Kenneth Keller

Director, Arts & Sciences Computing

Co-director, Humanities Digital Workshop

Doug Knox

Assistant Director, Humanities Digital Workshop

Stephen Pentecost

Digital Humanities Specialist

Michael Dango

Post Baccalaureate Fellow

Anupam Basu

Postdoctoral Fellow (starting in July, 2012)

Various (and numerous) faculty, graduate students, and undergraduates.

Projects in the Humanities Digital Workshop?



Bizet
Thematic
Catalog



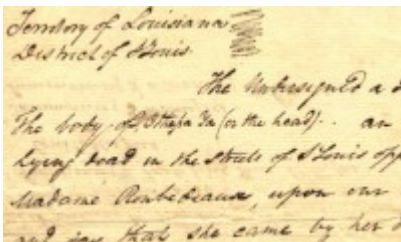
Rethinking the History of
German Literature 1731-1864:
A Statistical Approach



Creating a
Federal
Government



The Spenser Archive



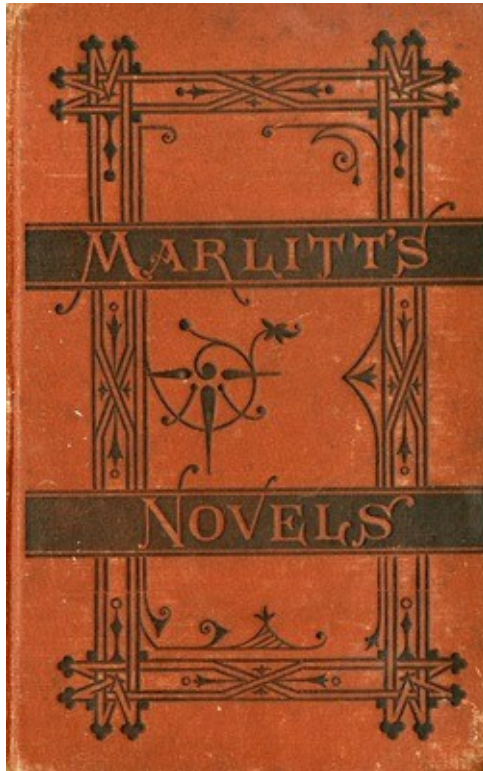
St. Louis
Circuit Court
Historical
Records



The American Publication
History of 19th Century
German Novels

And a few more . . .

The American Publication History of 19th Century German Novels



German-American Novels of the 19th Century - Google Chrome

German-American Novels x

talus.artsci.wustl.edu:8881/listPublicationHistoryViewOnlyNew

| | | | |
|------|--------------------|---|--------|
| 10 | publication | The Old Man's Secret (Wister) 1868 | change |
| 100 | publication | The Old Maid's Secret (Unknown) 1882 | change |
| 1001 | publication | Higher than the Church, an art legend of ancient times, by Wilhelmine von Hillern 1881 | change |
| | city | New York | |
| | date | 1881 | |
| | hasPublisher | | |
| | title | Higher than the Church, an art legend of ancient times, by Wilhelmine von Hillern | |
| | hasPublication rev | | |
| 1002 | publication | Um hohen Preis 1878 | change |
| 1007 | publication | Higher than the Church, by Wilhelmine von Hillern 1916 | change |
| 1009 | publication | Higher than the Church. 1902 | change |
| 1011 | publication | Alpenfee 1888 | change |
| 1012 | publication | The Alpine Fay 1889 | change |

Zora Clevenger

Checked out 57 books

Horatio Alger (17)

Charles Austin Fosdick (10)

Edward Ellis (8)

William T Adams (6)

Plus Cowpers, and a couple of items from the Congress.



Wayman Adams

Checked out 152 books

Charles Austin Fosdick (24)

Edward Ellis (15)

Horatio Alger (9)

William T Adams (5)

Plus Milton, and a couple of items from Congress.

Plus art, Italian, The Edinburgh review, etc



What did we do to the data?

1. **We asked Jim Connolly for help.**
2. De-normalized (“flattened”) the data.
3. **Handled borrower-patron census questions.**

Borrower ↔ Patron ↔ Census information

Rev Lake Woodard ↔ Kate Wilson ↔ White female born in 1836

178,000 transactions reduced to 108,000

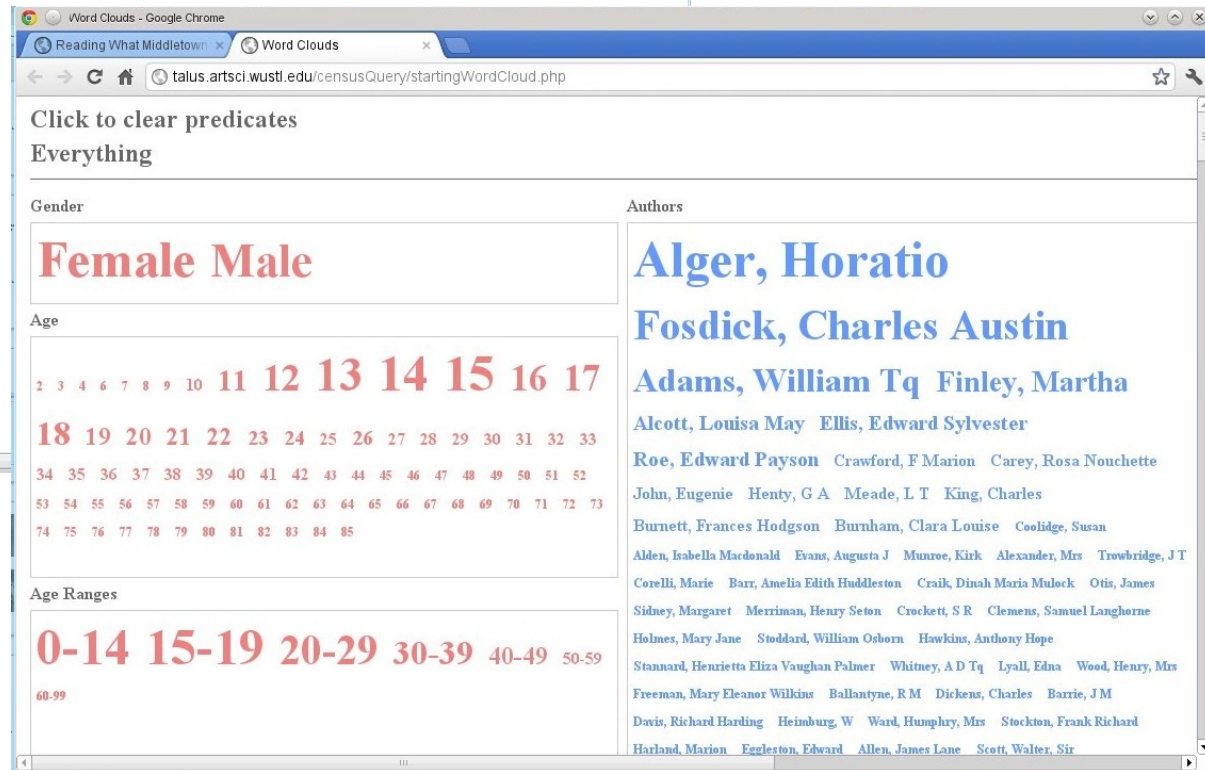
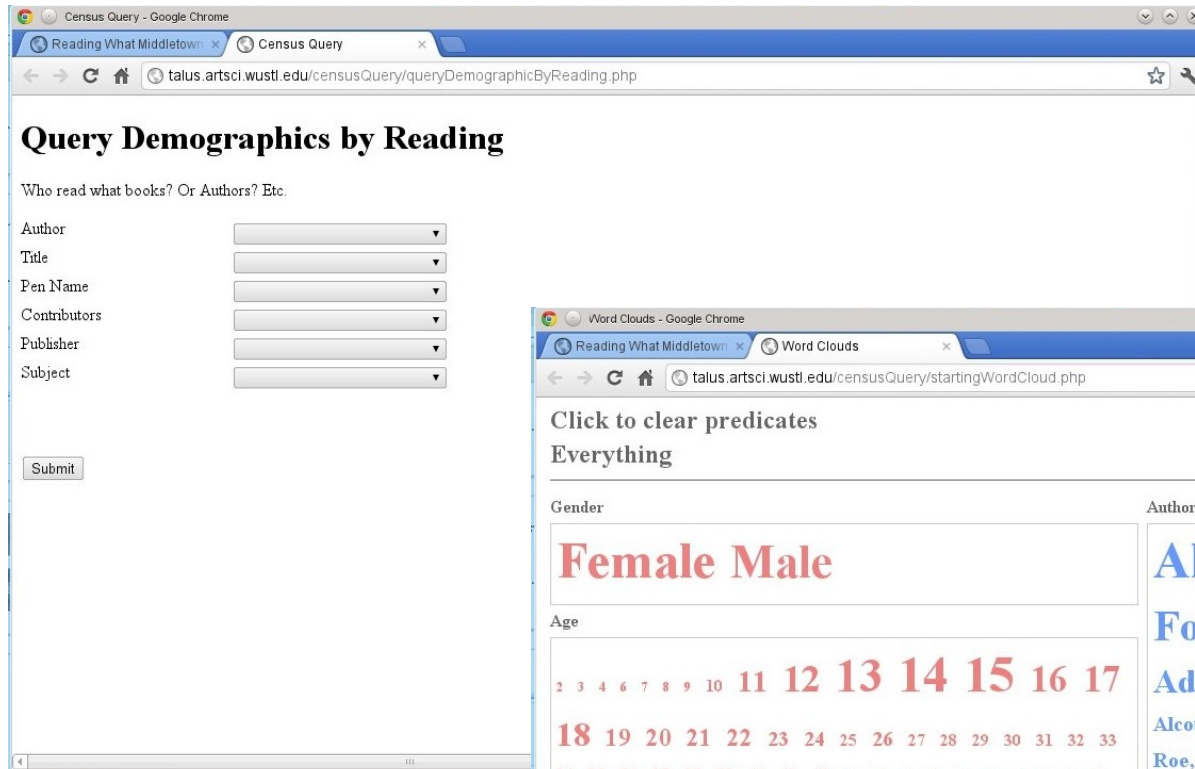
4. **Reduced the scale of the data** by focusing only on frequent borrowers, authors and titles before performing some kinds of analysis.

10,000 borrowers, 1,700 authors, 4,000 titles
29 billion evaluations for author overlaps
160 billion for titles

vs

1,000 borrowers, 480 authors, 1,200 titles
230 million for authors
1.4 billion for titles

Query interfaces

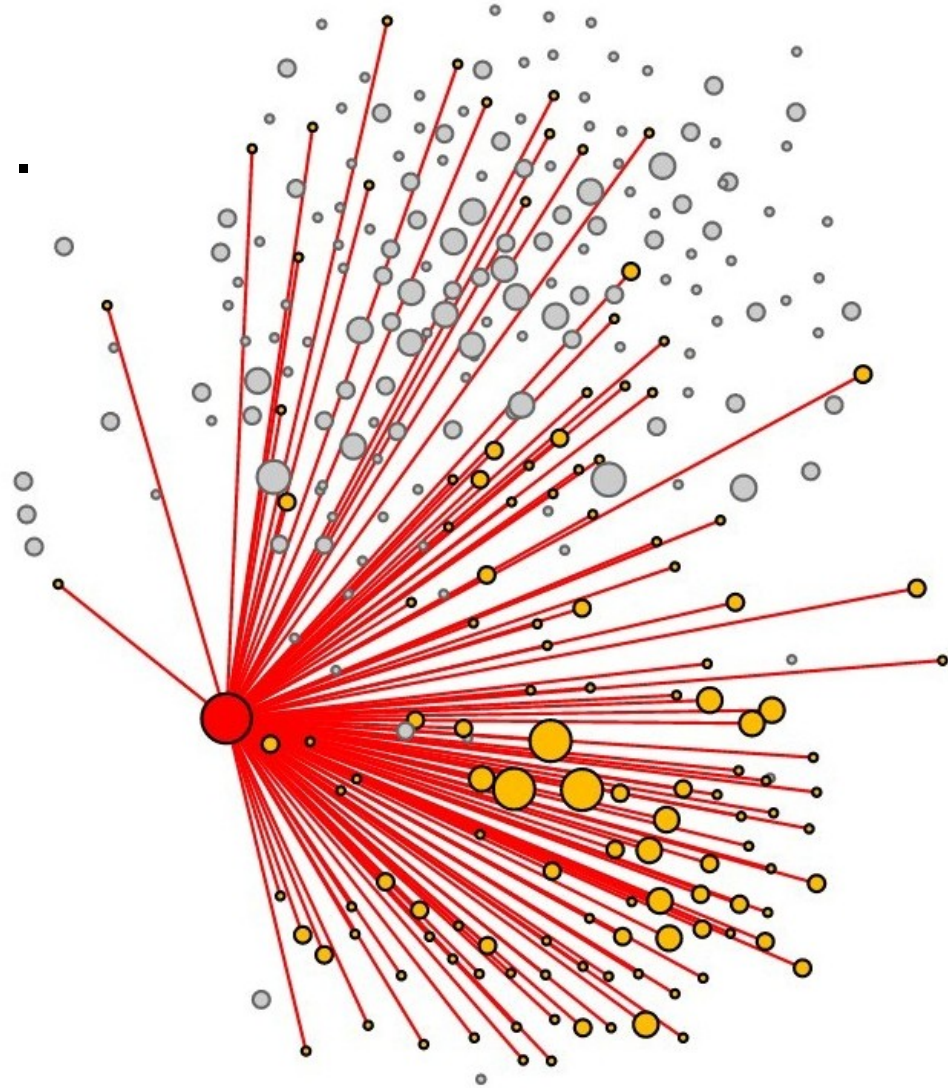


Market Basket Analysis

Amazon recommends . . .

or

Beer & diapers



(a) Typical Readers

Does market
basket analysis
overstate
common reading
patterns?

How different are
readers?

Men 14 years old and younger

Commonly read authors

Ellis, Edward Sylvester 197 readers (83%)

Alger, Horatio 218 readers (92%)

Trowbridge, J T 126 readers (53%)

Stoddard, William Osborn 113 readers (48%)

Munroe, Kirk 116 readers (49%)

Adams, William Tq 191 readers (81%)

Ballantyne, R M 110 readers (46%)

King, Charles 97 readers (41%)

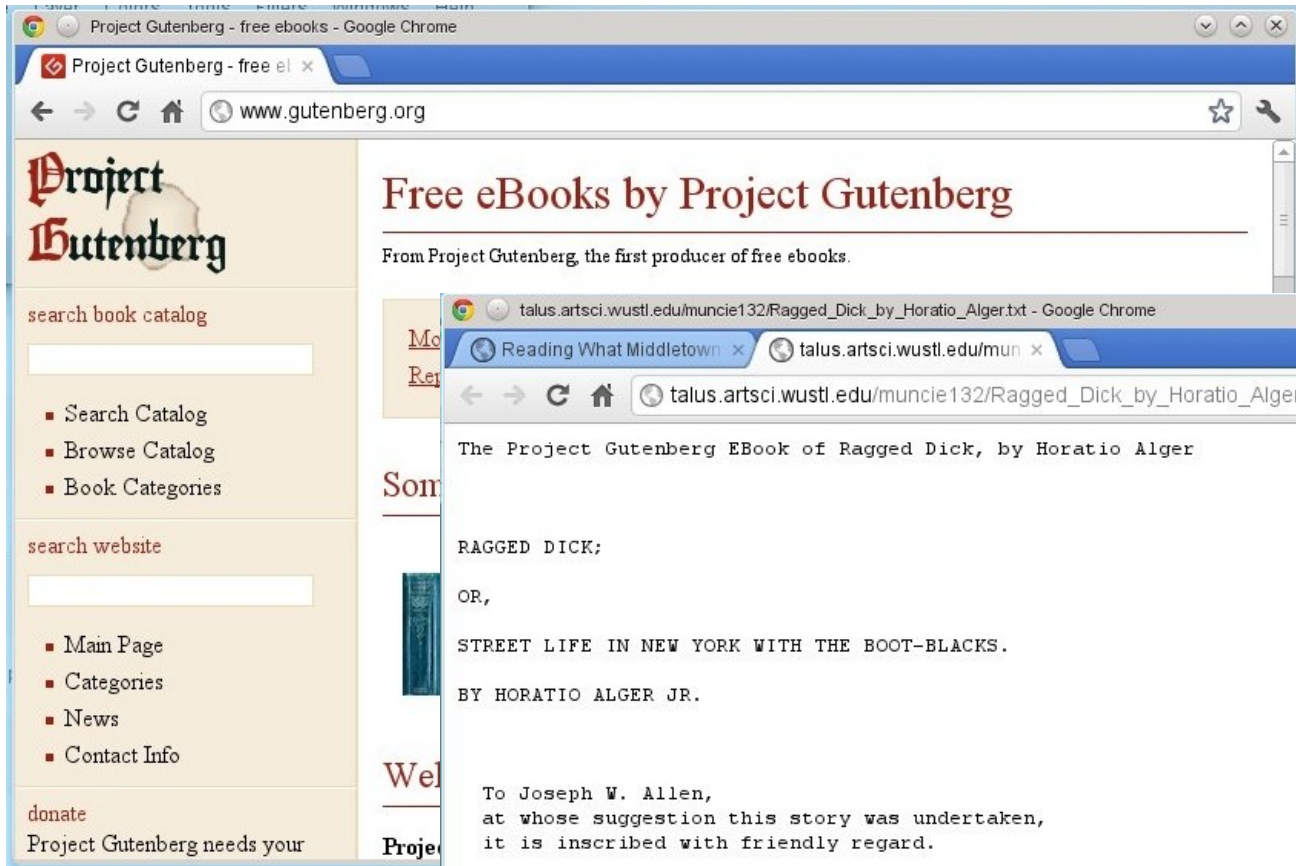
Alcott, Louisa May 106 readers (45%)

Henty, G A 109 readers (46%)

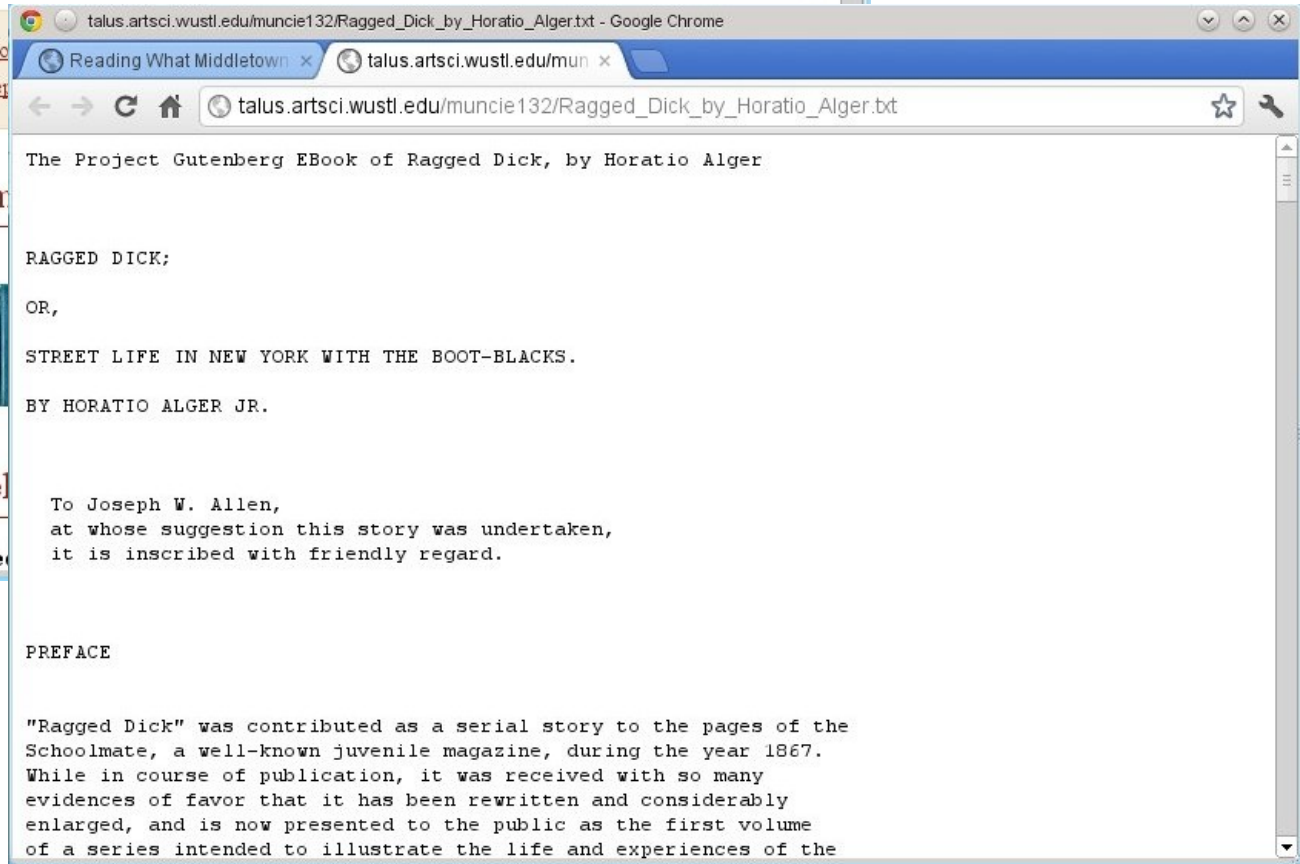
Fosdick, Charles Austin 214 readers (91%)

Otis, James 122 readers (51%)

Texts



The screenshot shows the Project Gutenberg website in a Google Chrome browser. The address bar displays 'www.gutenberg.org'. The page features the Project Gutenberg logo on the left, a search bar, and a navigation menu with links to 'Search Catalog', 'Browse Catalog', and 'Book Categories'. The main content area is titled 'Free eBooks by Project Gutenberg' and includes the text 'From Project Gutenberg, the first producer of free ebooks.'



The screenshot shows a text file viewer displaying the content of 'Ragged Dick' by Horatio Alger. The text is as follows:

The Project Gutenberg eBook of Ragged Dick, by Horatio Alger

RAGGED DICK;

OR,

STREET LIFE IN NEW YORK WITH THE BOOT-BLACKS.

BY HORATIO ALGER JR.

To Joseph W. Allen,
at whose suggestion this story was undertaken,
it is inscribed with friendly regard.

PREFACE

"Ragged Dick" was contributed as a serial story to the pages of the Schoolmate, a well-known juvenile magazine, during the year 1867. While in course of publication, it was received with so many evidences of favor that it has been rewritten and considerably enlarged, and is now presented to the public as the first volume of a series intended to illustrate the life and experiences of the

Reading Level

| Novel | Grade Level | Reading Ease | Sentences | Words | Syllables/Word | Words/Sentence |
|---|-------------|--------------|-----------|---------|----------------|----------------|
| Sam's Chance, by Horatio Alger | 4.39 | 81.95 | 2,629 | 27,089 | 1.35 | 10.30 |
| The Young Outlaw, by Horatio Alger | 4.60 | 81.31 | 2,746 | 29,668 | 1.35 | 10.80 |
| Wait and Hope, by Horatio Alger | 4.68 | 80.49 | 2,840 | 30,282 | 1.37 | 10.66 |
| Bound to Rise, by Horatio Alger | 4.96 | 80.08 | 2,793 | 32,253 | 1.36 | 11.55 |
| Herbert Carter's Legacy, by Horatio Alger | 5.02 | 78.78 | 2,720 | 30,070 | 1.38 | 11.06 |
| The Wide, Wide World, by Susan Warner | 5.04 | 82.15 | 15,875 | 206,926 | 1.32 | 13.03 |
| Adventures of Huckleberry Finn, Complete by Mark Twain (Samuel Clemens) | 5.09 | 86.20 | 6,058 | 93,972 | 1.24 | 15.51 |
| The Telegraph Boy, by Horatio Alger, Jr. | 5.12 | 78.83 | 2,359 | 27,147 | 1.37 | 11.51 |
| The Young Surveyor, by J. T. Trowbridge | 5.12 | 81.59 | 3,773 | 49,293 | 1.32 | 13.06 |
| Helping Himself, by Horatio Alger | 5.14 | 78.96 | 2,875 | 33,514 | 1.37 | 11.66 |
| The Young Explorer, by Horatio Alger | 5.14 | 79.07 | 2,471 | 28,957 | 1.37 | 11.72 |
| The Young Adventurer, by Horatio Alger | 5.24 | 77.52 | 2,882 | 32,366 | 1.39 | 11.23 |
| Slow and Sure, by Horatio Alger | 5.26 | 78.27 | 2,755 | 32,315 | 1.38 | 11.73 |
| The Prisoner of Zenda, by Anthony Hope | 5.26 | 81.65 | 3,340 | 45,608 | 1.32 | 13.66 |
| Jack's Ward, by Horatio Alger, Jr. | 5.31 | 77.16 | 2,894 | 32,746 | 1.40 | 11.32 |
| Louisiana, by Frances Hodgson Burnett | 5.32 | 80.07 | 1,997 | 25,953 | 1.34 | 13.00 |
| A Sweet Girl Graduate, by Mrs. L. T. Meade | 5.40 | 79.11 | 4,728 | 60,385 | 1.36 | 12.77 |
| Rupert of Hentzau, by Anthony Hope | 5.40 | 80.36 | 5,473 | 73,781 | 1.33 | 13.48 |
| Verner's Pride, by Mrs. Henry Wood | 5.43 | 78.01 | 17,896 | 219,616 | 1.38 | 12.27 |
| Polly, by L. T. Meade | 5.55 | 78.63 | 5,177 | 67,939 | 1.36 | 13.12 |
| Five Little Peppers Midway, by Margaret Sidney | 5.60 | 79.52 | 4,025 | 55,674 | 1.34 | 13.83 |
| Dora Thorne, by Charlotte M. Braeme | 5.74 | 77.69 | 6,835 | 91,301 | 1.37 | 13.36 |
| Five Little Peppers And How They Grew, by Margaret Sidney | 5.75 | 79.18 | 3,784 | 53,793 | 1.34 | 14.22 |
| Paul the Peddler, by Horatio Alger, Jr. | 5.76 | 76.27 | 2,504 | 21,740 | 1.20 | 12.68 |

Topic Modeling (topic percentages)

muncie132scrubbed_KS_RS_ES_enEarlyModern_50 (1).xls - LibreOffice Calc <@linux-v09m.site>

File Edit View Insert Format Tools Data Window Help

B14 Not_Like_Other_Girls_by_Rosa_N_Carey.txt

| | A | B | C | D | E | F | G | H | I |
|----|-----|--|---------|--------|---------|--------|---------|--------|---------|
| | | | Topic 0 | Pct 0 | Topic 1 | Pct 1 | Topic 2 | Pct 2 | Topic 3 |
| 1 | | | | | | | | | |
| 2 | 22 | Donald_and_Dorothy_by_Mary_Mapes_Dodge.txt | 0 | 15.76% | 29 | 12.41% | 12 | 11.20% | |
| 3 | 116 | Little_Lord_Fauntleroy_by_Frances_Hodgson_Burnett.txt | 0 | 48.97% | 42 | 6.08% | 29 | 5.54% | |
| 4 | 127 | A_Fair_Barbarian_by_Francis_Hodgson_Burnett.txt | 0 | 31.32% | 42 | 12.20% | 5 | 11.12% | |
| 5 | 1 | Five_Little_Peppers_Grown_Up_by_Margaret_Sidney.txt | 1 | 57.48% | 19 | 8.59% | 29 | 7.25% | |
| 6 | 13 | Polly_by_L_T_Meade.txt | 1 | 30.80% | 29 | 10.04% | 19 | 8.45% | |
| 7 | 49 | Five_Little_Peppers_Midway_by_Margaret_Sidney.txt | 1 | 49.24% | 29 | 10.40% | 19 | 9.57% | |
| 8 | 120 | Five_Little_Peppers_And_How_They_Grew_by_Margaret_Sidney.txt | 1 | 57.46% | 29 | 12.86% | 19 | 8.10% | |
| 9 | 26 | The_Prisoner_of_Zenda_by_Anthony_Hope.txt | 2 | 37.10% | 45 | 8.66% | 19 | 8.55% | |
| 10 | 41 | Janice_Meredith_by_Paul_Leicester_Ford.txt | 2 | 32.53% | 4 | 15.27% | 45 | 7.70% | |
| 11 | 109 | Rupert_of_Hertzau_by_Anthony_Hope.txt | 2 | 35.89% | 45 | 11.74% | 19 | 11.44% | |
| 12 | 0 | Uncle_Max_by_Rosa_Nouchette_Carey.txt | 3 | 37.91% | 42 | 11.02% | 29 | 8.49% | |
| 13 | 24 | Doctor_Luttrells_First_Patient_by_Rosa_Nouchette_Carey.txt | 3 | 32.40% | 42 | 10.44% | 29 | 7.67% | |
| 14 | 79 | Not_Like_Other_Girls_by_Rosa_N_Carey.txt | 3 | 23.18% | 42 | 13.91% | 29 | 9.73% | |
| 15 | 92 | Hugh_Wynne_Free_Quaker_by_S_Weir_Mitchell.txt | 4 | 20.49% | 2 | 16.83% | 42 | 7.75% | |
| 16 | 93 | Winning_His_Way_by_Charles_Carleton_Coffin.txt | 4 | 21.22% | 19 | 12.68% | 11 | 12.55% | |
| 17 | 119 | A_War_Time_Wooring_by_Charles_King.txt | 4 | 31.13% | 45 | 8.62% | 25 | 6.38% | |
| 18 | 98 | The_Story_of_an_Untold_Love_by_Paul_Leicester_Ford.txt | 5 | 11.92% | 48 | 11.32% | 42 | 11.06% | |
| 19 | 44 | The_Earth_Trembled_by_E_P_Roe.txt | 6 | 38.38% | 48 | 18.99% | 42 | 8.87% | |
| 20 | 100 | An_Original_Belle_by_E_P_Roe.txt | 6 | 24.63% | 48 | 22.84% | 4 | 11.13% | |

Sheet1

Find

Sheet 1 / 1 | PageStyle_Sheet1 | STD | Sum=0 | 125%

Topic Modeling (topic words)

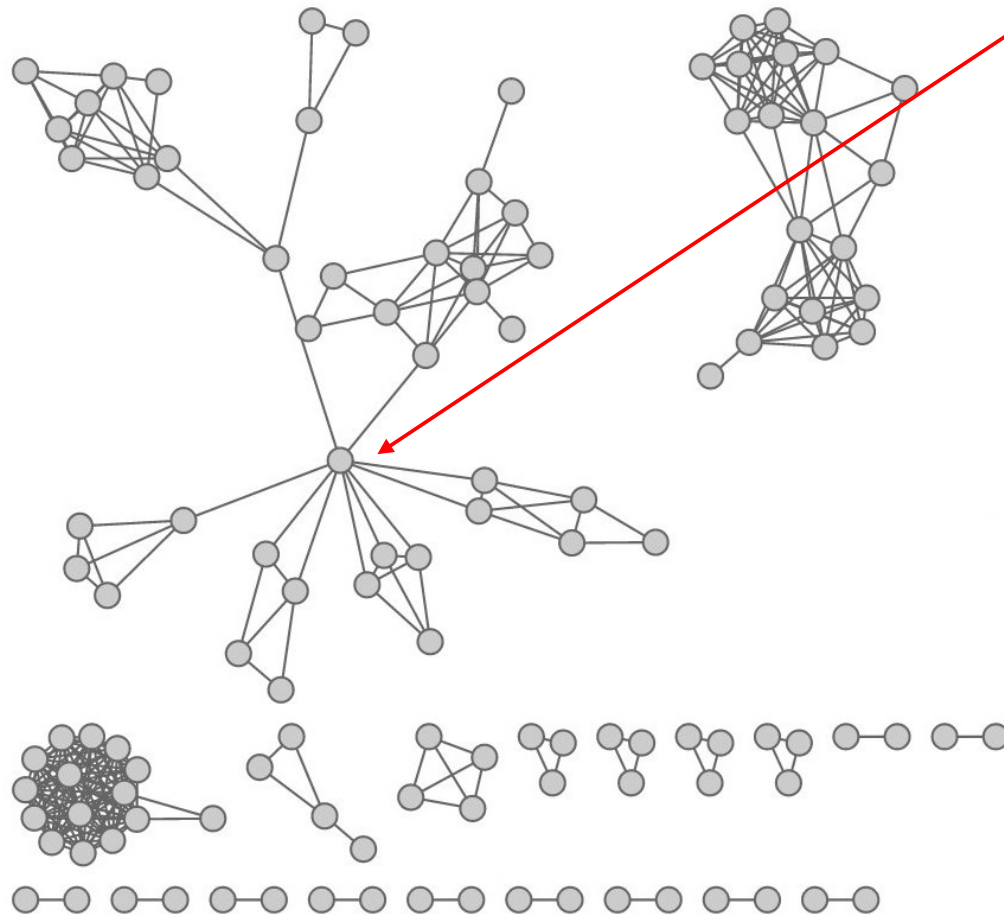


Topic Modeling (topic 4)

colonel army officers officer major fort camp soldiers soldier troops march british wounded guns regiment arthur st prisoners fighting sergeant lieutenant military governor saddle ranks fought virginia cavalry sword staff cannon government firing marched warren comrades england retreat uniform fled stable tent captured marching leader column artillery victory armed prisoner gallant fires batteries americans henry spy mount gun regiments prison fired pistol headquarters surrender roads battery commanded surgeon forces committee bridge halt mounted shots stream bullets resistance wound bridle powder brigade tents arnold guards issued band hunt gallop battles capture ships yards lee parade muskets frequent infantry armies foe

| | |
|--|--------------|
| A War Time Wooing by Charles King | 31.1% |
| Elsie at Viamede by Martha Finley | 22.6% |
| Winning His Way by Charles Carleton Coffin | 21.2% |
| Hugh Wynne Free Quaker by S Weir Mitchell | 20.5% |
| George at the Fort by Harry Castlemon | 18.8% |
| Elsies Vacation and After Events by Martha Finley | 18.2% |
| Rodney The Partisan by Harry Castlemon | 17.9% |
| Janice Meredith by Paul Leicester Ford | 15.3% |
| Frank on the Lower Mississippi by Castlemon | 12.5% |
| His Sombre Rivals by E P Roe | 11.2% |
| An Original Belle by E P Roe | 11.1% |
| Frank on a Gun Boat by Harry Castlemon | 11.0% |

Topic Modeling (distance)



The Story of A Bad Boy Thomas Bailey Aldrich

This is the story of a bad boy. Well, not such a very bad, but a pretty bad boy; and I ought to know, for I am, or rather I was, that boy myself.

Lest the title should mislead the reader, I hasten to assure him here that I have no dark confessions to make. I call my story the story of a bad boy, partly to distinguish myself from those faultless young gentlemen who generally figure in narratives of this kind, and partly because I really was not a cherub. I may truthfully say I was an amiable, impulsive lad, blessed with fine digestive powers, and no hypocrite. I didn't want to be an angel and with the angels stand; I didn't think the missionary tracts presented to me by the Rev. Wibird Hawkins were half so nice as Robinson Crusoe; and I didn't send my little pocket-money to the natives of the Feejee Islands, but spent it royally in peppermint-drops and taffy candy. In short, I was a real human boy, such as you may meet anywhere in New England, and no more like the impossible boy in a storybook than a sound orange is like one that has been sucked dry. But let us begin at the beginning.

Conclusions and questions

1. Thanks to Ball State University and the Muncie Public Library.
2. The HDW's work still needs revision, systematic review, etc.
3. What about the long tail? Unusual or infrequent reading?
4. Can we systematically discriminate between different kinds of reading (entertainment, educational, civic, occupational)?
5. How might we integrate data about reading (the *What Middletown Read* data with distant reading (topic modeling, etc)?
6. Can we generalize from the OCLC_subject information in the database?