NYU LIBRARIES

Infoloom

Topic maps for digital scholarly monographs

DC-2017 // October 28, 2017

Alexandra Provo
Project Manager & Digital Production Editor
Enhanced Networked Monographs Project
Digital Library Technology Services

alexandra.provo@nyu.edu / @AlexandraAlisa

Michel Biezunski
Innovator
Founder/CEO of Infoloom
Former Editor of the Topic Maps standard

mb@infoloom.com

Today's presentation

- Background on the Enhanced Networked Monographs project
- Topic maps introduction
- Topic Curation Toolkit software
- Topic map publishing
- Topic curation

Enhanced Networked Monographs (ENM)

- Funded by The Andrew. W. Mellon Foundation
- Free web access to ~110 selected EPUBs
  - NYU Press
  - University of Minnesota Press
  - University of Michigan Press

Enhanced Networked Monographs (ENM)

1. Platform for reading
2. Semantic text-manipulation tool (Topic Curation Toolkit)
3. Open-source tools and workflows
4. List of requirements

openaccessbooks.nyupress.org

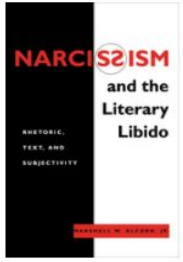
NYU PRESS

OPEN ACCESS BOOKS

Search for books...

NYU Press is pleased to offer these books in an open access platform for reading on desktop and mobile devices.
We will continue to contribute new titles each year. Read more about the project...

BROWSE ALL TITLES

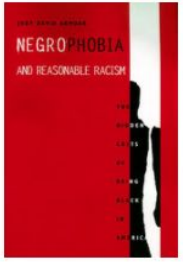


Narcissism and the Literary Libido

Rhetoric, Text, and Subjectivity

by Marshall W. Alcorn Jr.

What is it that makes language powerful? This book uses the psychoanalytic concepts of narcissism and libidinal investment to explain how rhetoric compels us and how it can effect change. The works of Joseph Conrad, James Baldwin, Michael Foucault, Jacques Derrida, Arthur Miller, D.H.



Negrophobia and Reasonable Racism

The Hidden Costs of Being Black in America

by Jody David Armour

Tackling the ugly secret of unconscious racism in American society, this book provides specific solutions to counter this entrenched phenomenon.

View Details | Read Online

Adventures of the Mind

The Memoirs of Natalie Clifford Barney

by Natalie Clifford Barney, John Spalding Gatton

In this book, Barney explores her family tree, chronicles her friendships and associations through reprinted correspondence and recreated conversations, and evokes the golden age of her salon in gallery of literary portraits.

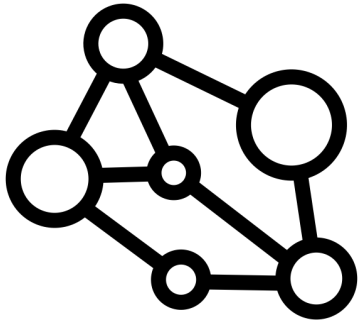
View Details | Read Online

About | Rights Information | NYU Press | Powered by NYU DLTS

How will we enhance and network monographs?

Created by Nikolay Necheuhin
from Noun Project

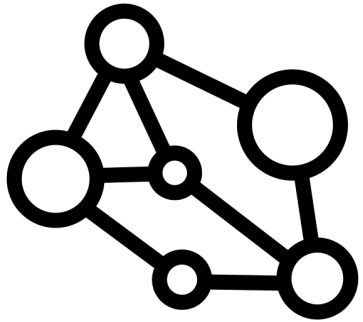
Created by Deepz
from Noun Project



Created by Adnen Kadri
from Noun Project

Today's focus

1. Platform for reading
2. **Semantic text-manipulation tool (Topic Curation Toolkit)**
3. Open-source tools and workflows
4. List of requirements



Created by Adnen Kadri
from Noun Project

Related work

Subject metadata and discovery

Nowick, E. A. ., & Stein, S. (2010). A Comparison of Term Clusters for Tokenized Words Collected from Controlled Vocabularies, User Keyword Searches, and Online Documents

Yang, L. (2016). Metadata Effectiveness in Internet Discovery: An Analysis of Digital Collection Metadata Elements and Internet Search Engine Keywords

Other projects

Meta-Dex (University of British Columbia)

Huggett, M., & Rasmussen, E. (2012). Dynamic Online Views of Meta-indexes

Topic Maps at a glance

Or, how to represent complex knowledge given the limitations of computer systems.

Formats

Computers need predefined formats to access information

Names

Computers need names to locate things.

Hierarchies

Navigation is done either through search or by navigating taxonomies.

“Topic” “Maps”

Topic Maps is an information architecture that provides access to meaning independently of its sources. (ISO/IEC 13250)

Topics

Unit of meaning occupying a location in an abstract space.

Any subject of conversation is a topic.

Maps

Graph providing the connection between the topics

Topic Names

Topics can have one name, multiple names, and no name.

- **Multiple names**
  - Synonyms, multiple languages, variant names for different audiences (technical vs. common, etc.)
- **No name**
  - Cross-references, web links are links between 2 occurrences of the same topic.
- **One name for multiple topics**
- **Names can be topics**
  - The naming process itself can come under discussion.

The Topic Curation Toolkit (TCT)

Features

- Ingest Ebooks
- Manage Topic Map
- Export to Reading Interface

Technology

- Backend: Django / PostgresQL
- Frontend: AngularJS

TCT Ingest

Every entry and subentry in the indexes become a topic.

Automatic relationships created:

- Entries/Subentries
- See / See also
- Containment
- 3 word rule

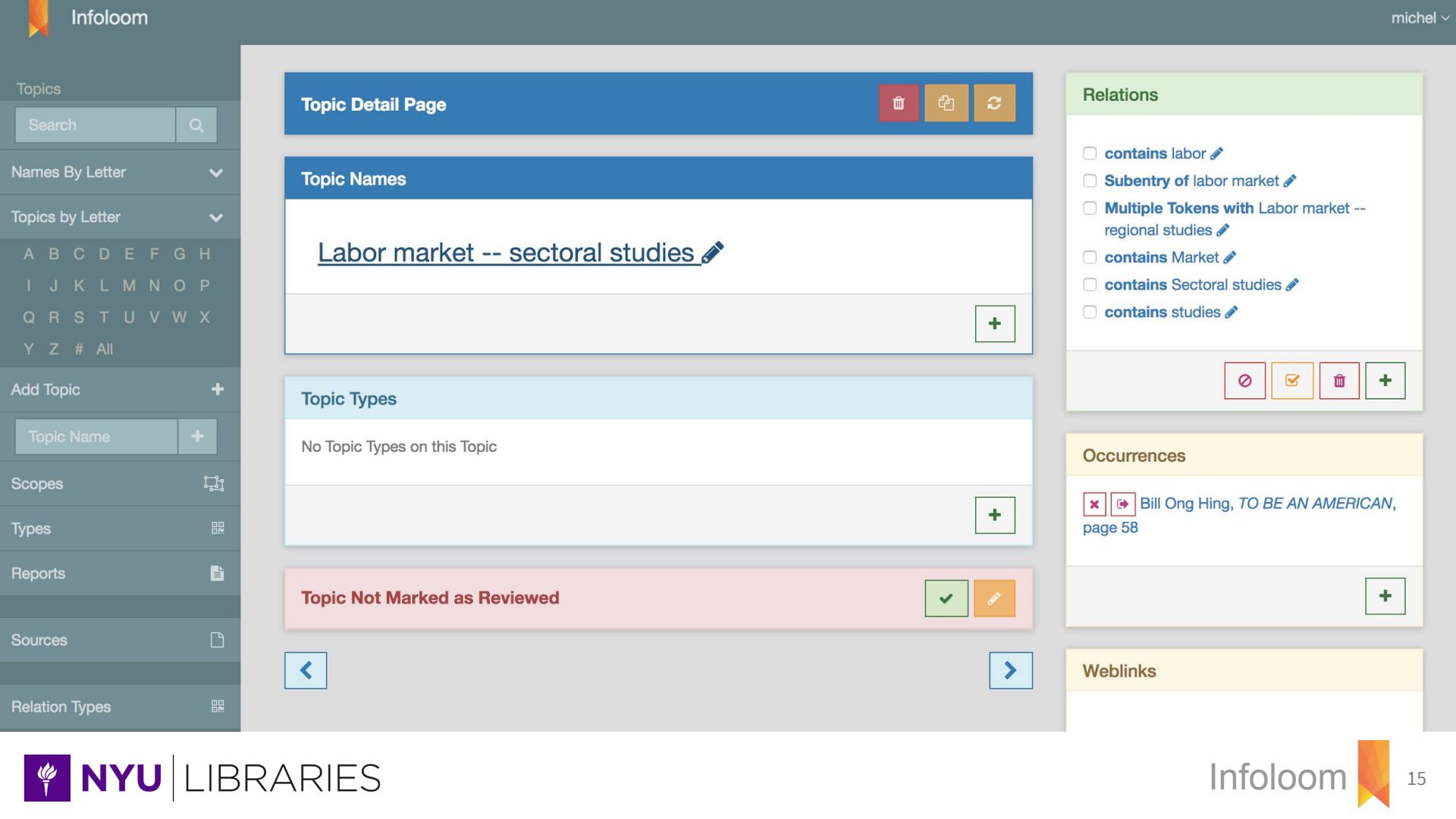
TCT Editorial Interface

Enables navigation between topics and pages in the books

Look up topics with a master index and topic search

Editing interface to add, delete or modify:

- Names, types, occurrences, sources, relations, weblinks


Infoloom
michel
Topics
Search
Names By Letter
Topics by Letter
A B C D E F G H
I J K L M N O P
Q R S T U V W X
Y Z # All
Add Topic
Topic Name
Scopes
Types
Reports
Sources
Relation Types
Topic Detail Page
Topic Names
Labor market -- sectoral studies
Topic Types
No Topic Types on this Topic
Topic Not Marked as Reviewed
Relations
contains labor
Subentry of labor market
Multiple Tokens with Labor market -- regional studies
contains Market
contains Sectoral studies
contains studies
Occurrences
Bill Ong Hing, TO BE AN AMERICAN, page 58
Weblinks

Content

that they are in some sense “employees.” And if administrators and faculty at Yale or elsewhere want to claim that their graduate students’ wages are not “wages” because their teaching (which is not strictly “teaching”) is merely part of their professional training as apprentice professors, then it makes sense to call the bluff: take graduate students out of the classrooms in which they work as graders, assistants, and instructors; maintain their stipend support at its current levels; and give them professional development and training that does not involve the direct supervision of undergraduates. Then we’ll see how long Yale University can survive without the labor (which is not strictly “labor”) of its graduate student teaching assistants. At the time, I thought my support for graduate student unions—in a speech delivered to, among other people, unionized graduate students—amounted to endorsing candidates after they’d won their elections. To my surprise, however, I learned later that the graduate students were very pleased with my speech, and that some even considered it “courageous.” It seems that I had denounced as ridiculous Yale administrators’ claims that graduate students were not employees in front of a number of Kansas administrators who had claimed that graduate students were not employees. (I told the students I had had no idea that my audience included actual bad faith negotiators, and that my “courage” in

Topics at This Location

- Faculty -- and graduate student unionization
- Graduate Employees and Students Organization (GESO), Yale -- faculty response to
- Unionization of graduate students -- opposition to
- University of Kansas
- Yale University anti-union stance of -- faculty response to graduate student unionization at

Indexes

Index 1

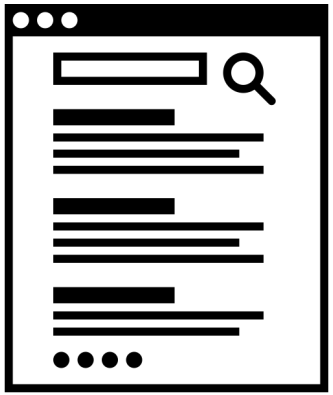
TCT Workflow Management

Topics marked as reviewed or modified

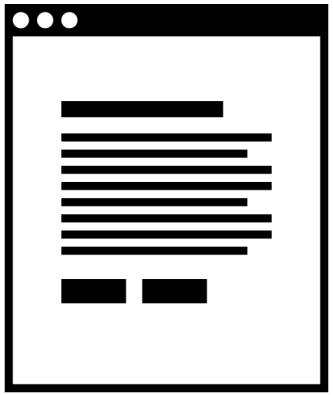
Batch process for data enrichment using existing standard classification schemas

Reports

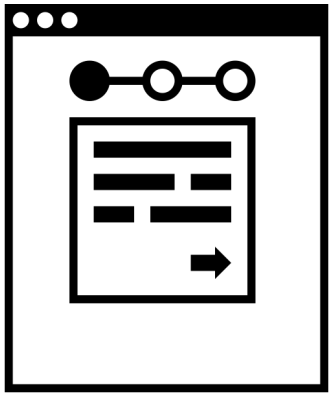
Topic map publishing



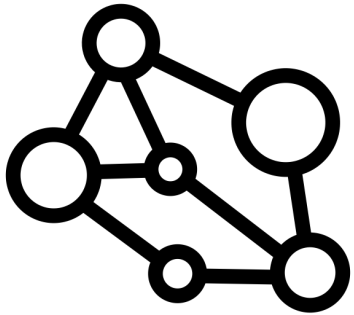
Created by Rflor
from Noun Project



Created by Rflor
from Noun Project



Created by Rflor
from Noun Project



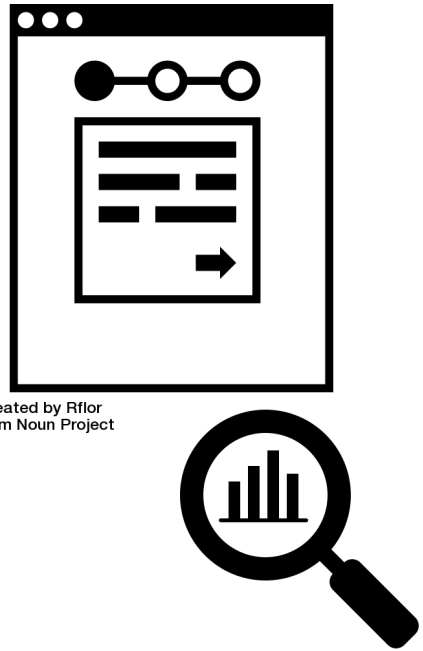
Created by Adnen Kadri
from Noun Project

Topic map publishing


```
"basket": {
        "id": 7399,
        "topic_hits": [
            {
                "id": 10210,
                "name": "Video game",
                "scope": {
                    "id": 2,
                    "scope": "Generic"
                },
                "bypass": false,
                "hidden": false,
                "preferred": false
            },
            {
                "id": 8758,
                "name": "video -- \"games\"",
                "scope": {
                    "id": 2,
                    "scope": "Generic"
                },
                "bypass": false,
                "hidden": false,
                "preferred": false
            }
        ],
        "occurs": [
            {
                "id": 28663,
                "location": {
                    "id": 6418,
                    "document": {
                        "title": "Making News at The New York Times",
                        "author": "Usher, Nikki"
                    },
                    "localid": "page_150"
                }
            },
```

Topic map publishing

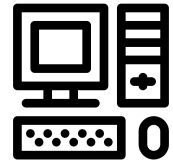
```
"@id" : "http://nyutct.org/topic/7399",
"@type" : "skos:Concept",
"skosxl:prefLabel" :       { "@id" : "http://nyutct.org/name/10210",
                             "skosxl:literalForm" : "Video Game",
                             "gvp:qualifier" : "Generic"
                           },

"skosxl:altLabel" : { "@id" : "http://nyutct.org/name/8758",
                             "skosxl:literalForm" : "video -- games",
                             "gvp:qualifier" : "Generic"
                           },
"skos:related" : [ http://nyutct.org/topic/209, http://nyutct.org/topic/90 ],

"skos:exactMatch": ["https://www.wikidata.org/wiki/Q7889"]
{
      "@id" : "http://nyutct.org/occurrence/31986",
      "@type" : "oa:Annotation",
      "oa:motivatedBy" : "oa:tagging",
      "oa:hasBody" :       { "@id" : "http://nyutct.org/topic/7399" },
      "oa:hasTarget" :     {
                             "oa:hasSource" : { "@id" : "BOOK ID (isbn? Some other unique ID?)"}
                             "oa:hasSelector" : CFI? Page # tag?
      }
}
```



Created by Rflor
from Noun Project

Created by Setyo Ari Wibowo
from Noun Project

Topic curation workflows



Created by Arthur Shlain
from Noun Project

Topic Curation Toolkit *automatically*:

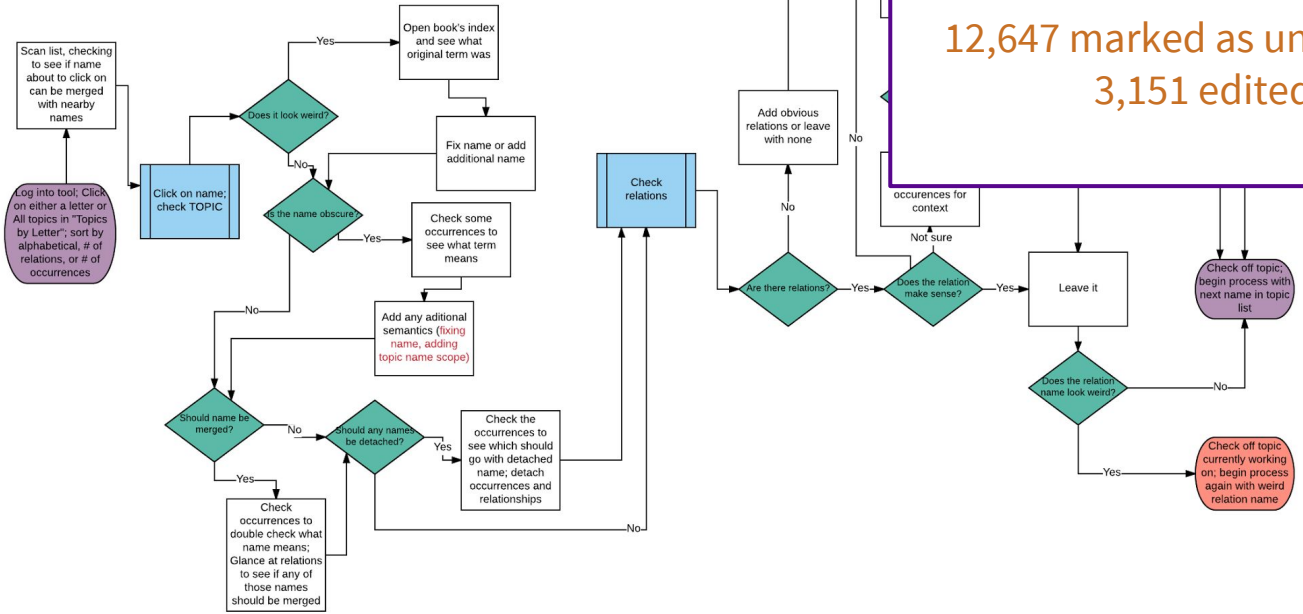
- merges topics
- creates relations between topics

Created by Adrien Coquet
from Noun Project

Human editor *manually*:

- fixes processing errors
- removes certain relations
- merges synonyms and splitting up homonyms
- adds relations tool did not find
- adding new data, sometimes from external sources like Wikidata

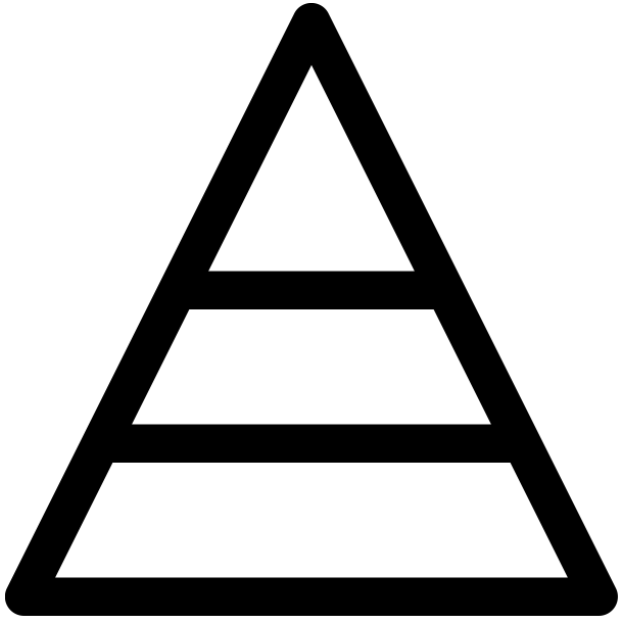
Topic curation workflows



15,800 topics checked
~45,000 topics from 88 indexes

12,647 marked as unchanged
3,151 edited

Topic curation workflows



Created by Icondesk
from Noun Project

Values

Principles

Practices

- adding and removing relations
- splitting, deleting, and merging topics
- adding scopes to names
- dealing with alternate names
- curating subentry relationships

Topic curation workflows

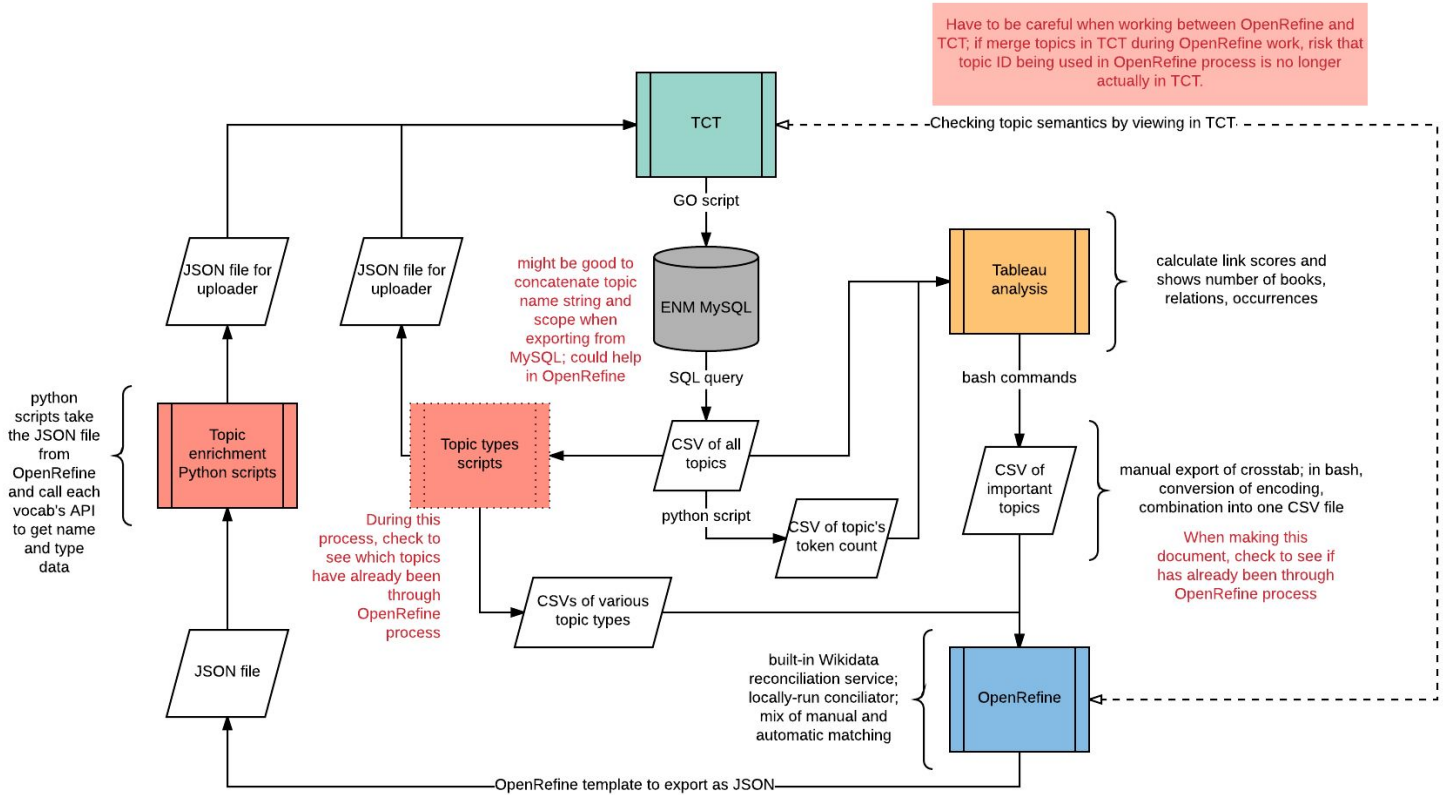


Table showing topic completeness and link scores

Tct Id	Topic Name	Completeness	Completeness (weighted)	Number of relations	Number of occurrences	Number of books	Number of words in top..	Link score (relations)	Link score (occurrences)	Link score (books)	Overall link score (ave..
14034	Freud, Sigmund	1.00	1.00	203	105	22	2	0.35	0.89	0.79	67%
9640	Foucault, Michel	1.00	1.00	11	73	28	2	0.02	0.62	1.00	55%
11928	United States of America	1.00	1.00	468	66	7	4	0.80	0.56	0.25	54%
22256	Woman	1.00	1.00	585	52	3	1	1.00	0.44	0.11	52%
7672	feminism	1.00	1.00	101	118	8	1	0.17	1.00	0.29	49%
826	The New York Times	1.00	1.00	35	89	18	4	0.06	0.75	0.64	49%
26141	Clinton, William Jefferson	1.00	1.00	259	61	12	3	0.44	0.52	0.43	46%
7907	race	1.00	1.00	165	99	7	1	0.28	0.84	0.25	46%
62	culture(s)	1.00	1.00	224	56	8	1	0.38	0.47	0.29	38%
11204	gender	1.00	1.00	146	63	9	1	0.25	0.53	0.32	37%
250	internet	1.00	1.00	21	105	5	1	0.04	0.89	0.18	37%
11796	self	1.00	1.00	510	10	4	1	0.87	0.08	0.14	37%
278	Google	1.00	1.00	26	85	9	1	0.04	0.72	0.32	36%
277	globalization	1.00	1.00	25	97	6	1	0.04	0.82	0.21	36%
3131	World War II	1.00	1.00	25	59	15	3	0.04	0.50	0.54	36%
6789	sexuality	1.00	1.00	100	69	9	1	0.17	0.58	0.32	36%
11138	Families	1.00	1.00	299	36	6	1	0.51	0.31	0.21	34%
17941	Kant -- Immanuel	1.00	1.00	13	67	12	2	0.02	0.57	0.43	34%
871	racism	1.00	1.00	98	62	9	1	0.17	0.53	0.32	34%
13522	power	1.00	1.00	196	57	5	1	0.34	0.48	0.18	33%
2822	King, Martin Luther, Jr.	1.00	1.00	19	44	16	4	0.03	0.37	0.57	33%
4071	Reagan, President Ronald	1.00	1.00	50	50	13	3	0.09	0.42	0.46	32%
11728	School(s)	1.00	1.00	408	18	3	1	0.70	0.15	0.11	32%
8541	values	1.00	1.00	83	74	5	1	0.14	0.63	0.18	32%
26738	American	1.00	1.00	501	4	1	1	0.86	0.03	0.04	31%
18416	Law, the	1.00	1.00	360	15	5	2	0.62	0.13	0.18	31%
7777	state, the	1.00	1.00	368	26	2	2	0.63	0.22	0.07	31%

Overall link score

0% 67%

Overall link score

0% 67%

Number of occurrences

0 118

Number of distinct books in which topic app...

0 28

Number of relations

0 585

Number of words in topic name

1 26

Metric based on Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories.

OpenRefine important_topics_to_enrich_July7 Permalink

Open... Export Help

Facet / Filter | Undo / Redo 163

Refresh | Reset All | Remove All

WikidataRecon: judgment change

3 choices Sort by: **name** count

matched 993
new 1
none 3499

Facet by choice counts

WikidataRecon: best candidate's types match? change

2 choices Sort by: **name** count

(unreconciled) 1192
false 3301

Facet by choice counts

WikidataRecon: best candidate's types change

542 choices Sort by: **name** count

(unreconciled) 1192
Q1002697 10
Q1007311 1
Q101352 12
Q1021128 1
Q10417670 1
Q1047113 1
Q1057504 1
Q1058914 1
Q1060829 1
Q1061197 1
Q106259 1

4493 rows Extensions:

Show as: **rows** records Show: 5 10 25 **50** rows « first ‹ previous **1 - 50** next › last »

All		Tct Id	Tct_link	Topic Name	WikidataRecon	Completeness	Completeness (v	Numb
1.	381	http://nyu.infoloom.nyc/basket/381	Addams, Jane	Jane Addams Choose new match	0.67	0.52	0
2.	467	http://nyu.infoloom.nyc/basket/467	Chicago Reader	Chicago Reader Choose new match	1.00	1.00	1
3.	595	http://nyu.infoloom.nyc/basket/595	Harvard University - - Kennedy School of Government	Harvard Kennedy School Choose new match	1.00	1.00	1
4.	735	http://nyu.infoloom.nyc/basket/735	Lupton, Julia Reinhardt	Lupton, Julia Reinhardt Create new topic Search for match	0.67	0.52	0
5.	931	http://nyu.infoloom.nyc/basket/931	Austin, Texas	Austin Choose new match	1.00	1.00	1
6.	1004	http://nyu.infoloom.nyc/basket/1004	Williams, Tennessee	Tennessee Williams Choose new match	1.00	1.00	1
7.	1017	http://nyu.infoloom.nyc/basket/1017	Yúdice, George	Yúdice, George Create new topic Search for match	0.67	0.52	0
8.	1034	http://nyu.infoloom.nyc/basket/1034	Asahi Shimbun	Asahi Shimbun Asahi Shimbun (100) Asahi Shimbun Company (90) Asahi Shimbun Digital (76) Asahi Shimbun Building (74) Asahi Shimbun Publications (67) ヘラルド朝日 (64) Asahi Shimbun Fukuoka Office (63) Asahi Shimbun Hokkaido Office (62) Asahi Shimbun Tokyo Head Office (59) Asahi Shimbun Osaka Head Office (59) Asahi Shimbun Seibu Head Office (59) Asahi Shimbun Nagoya Head Office (58) Kamikaze (23) Create new topic Search for match	0.67	0.52	0
9.	1049	http://nyu.infoloom.nyc/basket/1049	Batman	Batman Batman (100) Batman in film (100)	0.67	0.52	0

Subject metadata and topics

Subject headings

Find an item with a known subject

Find what a library has on a subject

Back-of-book indexes

Quick and easy access to information

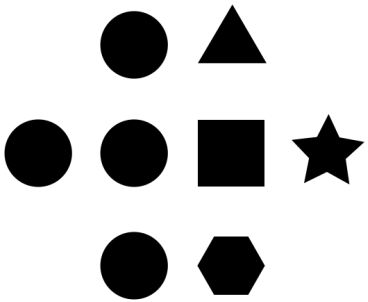
Sometimes used to give an overview of the text

item / place in text
collection / item

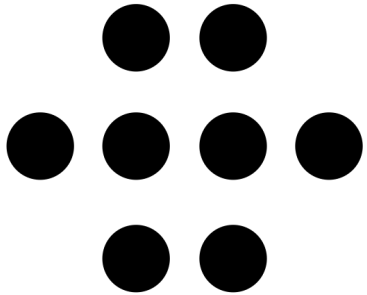
Nonscalability

“Against Cleaning” (Rawson & Muñoz, Curating Menus)

Nonscalability theory (Anna Tsing)



Created by dʒɛrmi Good
from Noun Project



Created by dʒɛrmi Good
from Noun Project

Specificity, exhaustivity, and granularity

- LCSH is specific but NOT exhaustive (apply 3-5 headings)
- Back-of-book-indexes are specific AND exhaustive
- Implicit meaning
    - **core group in networks and social worlds -- pair or small-group bonds via proximity in neighborhood**
    - *The Class: Living and Learning in the Digital Age*: 1,263 topics

Topics as subject headings?

Conclusion

References

Chan, L. M. (1990). *Library of Congress subject headings :principles of structure and policies for application /* (Annotated version.). Washington, D.C. : Retrieved from http://hdl.handle.net/2027/mdp.39015057586136

Coe, M. (2015). What do readers expect from book indexes and how do they use them? An exploratory user study. *Indexer*, *33*(3), 90–101.

Huggett, M., & Rasmussen, E. (2012). Dynamic Online Views of Meta-indexes. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 233–236). New York, NY, USA: ACM. https://doi.org/10.1145/2232817.2232860

Leise, F. (2016). Pan-granularism and specificity. *Indexer*, *34*(4), 147–155.

Mulvany, N. C. (2005). *Indexing Books, Second Edition* (2 edition). Chicago: University Of Chicago Press.

Nowick, E. A. ., & Stein, S. (2010). A Comparison of Term Clusters for Tokenized Words Collected from Controlled Vocabularies, User Keyword Searches, and Online Documents. Library Philosophy & Practice, 1–7.

References

Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries*, *10*(2–3), 67. https://doi.org/10.1007/s00799-009-0054-4

Rawson, K., & Muñoz, T. (2016). Against Cleaning. *Curating Menus*. Retrieved from http://www.curatingmenus.org/articles/against-cleaning/

Stauber, D. M. (2004). *Facing the text: content and structure in book indexing*. Eugene, Or.: Cedar Row Press.

Yang, L. (2016). Metadata Effectiveness in Internet Discovery: An Analysis of Digital Collection Metadata Elements and Internet Search Engine Keywords | Yang | College & Research Libraries. College & Research Libraries, 77(1). https://doi.org/https://doi.org/10.5860/crl.77.1.7

Thank you!

Contact us

Alexandra Provo
alexandra.provo@nyu.edu
Twitter: @AlexandraAlisa
https://wp.nyu.edu/enmproject

Michel Biezunski
mb@infoloom.com
https://infoloom.com/