Historical Text Datafication and Loss: Computational Recovery of Typographical Layout Logic on an RDF Graph Featuring ML Methods

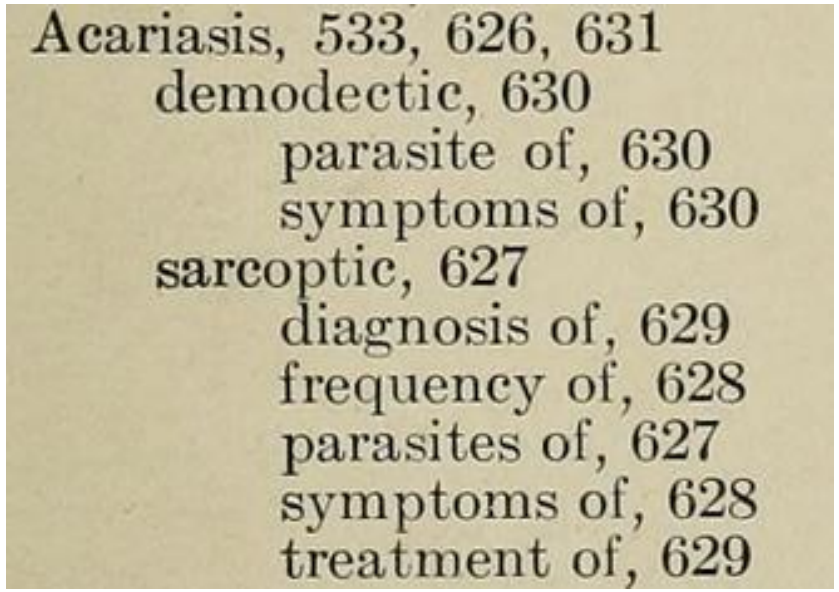
ASIS&T SIG AI Workshop | October 29, 2022

Huapu Liu
University of Alabama, USA - hliu68@crimson.ua.edu

Steven L. MacCall, PhD
University of Alabama, USA - smaccall@ua.edu

Linked Data Research Group: https://wikibase.slis.ua.edu/

Text Datafication and Loss of Typographical Layout Logic (TLL): Illustration of Problem


Acariasis, 533, 626, 631
demodectic, 630
parasite of, 630
symptoms of, 630
sarcoptic, 627
diagnosis of, 629
frequency of, 628
parasites of, 627
symptoms of, 628
treatment of, 629

Portion of book* index page scan for the index main entry “Acariasis”

OCR results for “Acariasis”

Acariasis, 533, 626, 631
demodectic, 630

parasite of, 630
symptoms of, 630
sarcoptic, 627

diagnosis of, 629
frequency of, 628
parasites of, 627
symptoms of, 628
treatment of, 629

*Osler, W., McCrae, T. (1907). Modern medicine: its theory and practice, in original contributions by American and foreign authors. Philadelphia: Lea brothers & co. Vol 1. https://hdl.handle.net/2027/nnc2.ark:/13960/t54f2j58c

Theoretical Framework: Navigational Paratexts

- Gerard Genette (1997) introduced concept of paratexts as functional agents that can “mediate the relations between text and reader.”
- Birke and Christ (2013) note that one important category of paratextual element is “navigational paratext:”
  - Such elements guide a reader’s reception in a more mechanical sense
  - Useful in approaching a text and for orienting a reader within the text
- Book index as navigational paratext:
  - Result of closed system indexing in contrast to open system (Klement, 2002)
  - List of entities mentioned within a single book along with locators (e.g., page #s)
  - Provide readers with granular navigation options within texts

Problem Description and Research Question

- For the Google Books Initiative, Google scanned and OCR'd every page of each book using same method.
- As a result, scanned and OCR'd book index pages have index entries without their indentations, which represents TLL that reduces intra-textual navigation compared to original print index.
- The restoration of an index's indentations would lead to testable questions in future research contexts (philology graphs).
- RQ: Can we recover TLL as logical hierarchies and represent them on a RDF knowledge graph?

High-level Workflow

1. Rescan book index pages from HathiTrust copies.
2. Connect hierarchies below index main entries with their respective subentries using partonomic logic property (“partOf”).
3. Load logical relationships on RDF knowledge graph (local Wikibase instance hosted by UA SLIS).
4. Convert page number locators to URLs for each scanned page and align with restored index entries on RDF knowledge graph.
5. Use SPARQL query to reconstruct an overlay HTML index for readers to return indexing navigation capabilities to book.

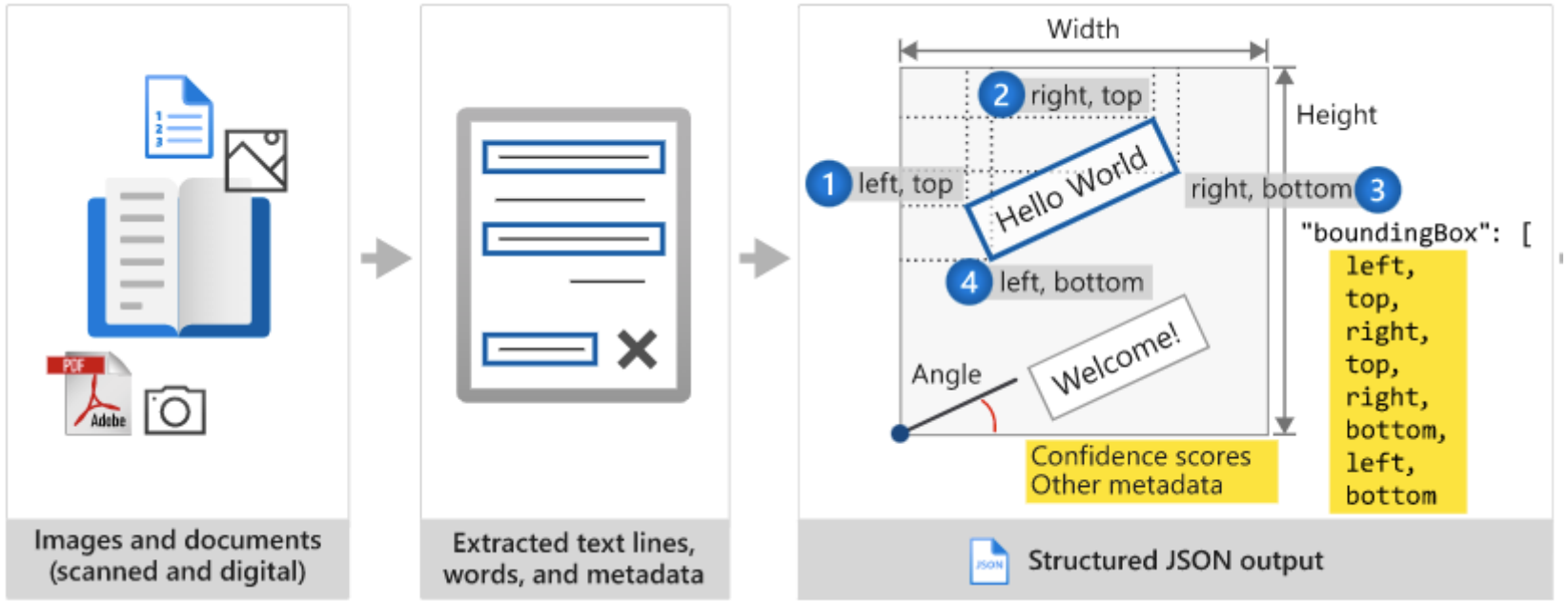
Scaling our Typographical Logic Workflow

- To scale our overlay index workflow, we developed a computational pipeline for detecting index indentations.
- Over the remaining slides, we will introduce some of the technical challenges and will describe the pipeline we developed to detect index indentations using cluster-based machine learning algorithms.

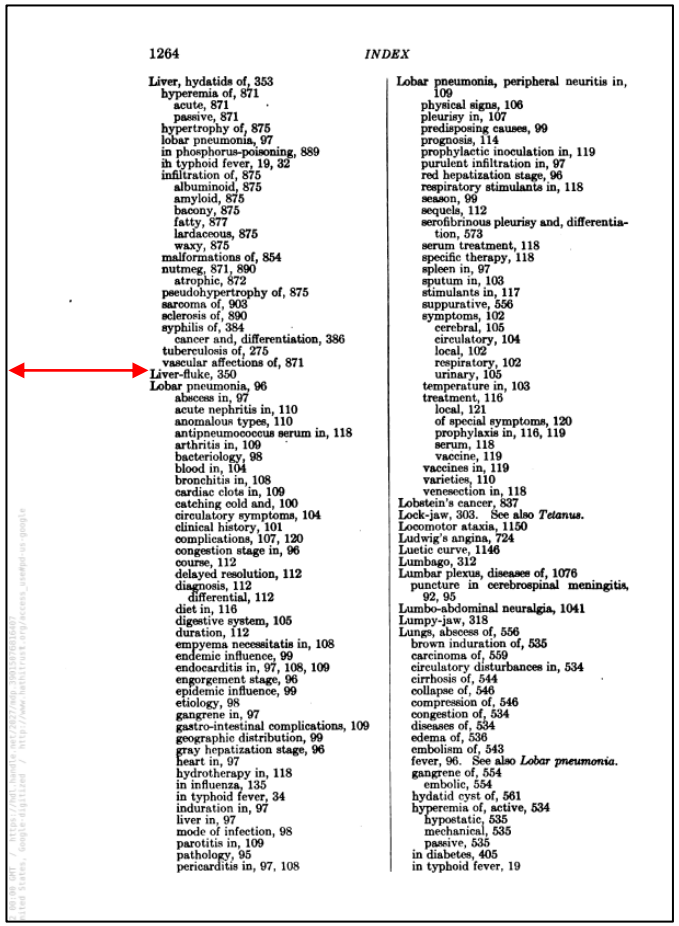
Computational Detection of Index Indentations

- Physical Structure Detection
- Technical Challenges
- Typographical Logic Recovery Pipeline
  - Noise Elimination
  - Column Segmentation
  - Skew Correction
  - Entry Classification
- Evaluation

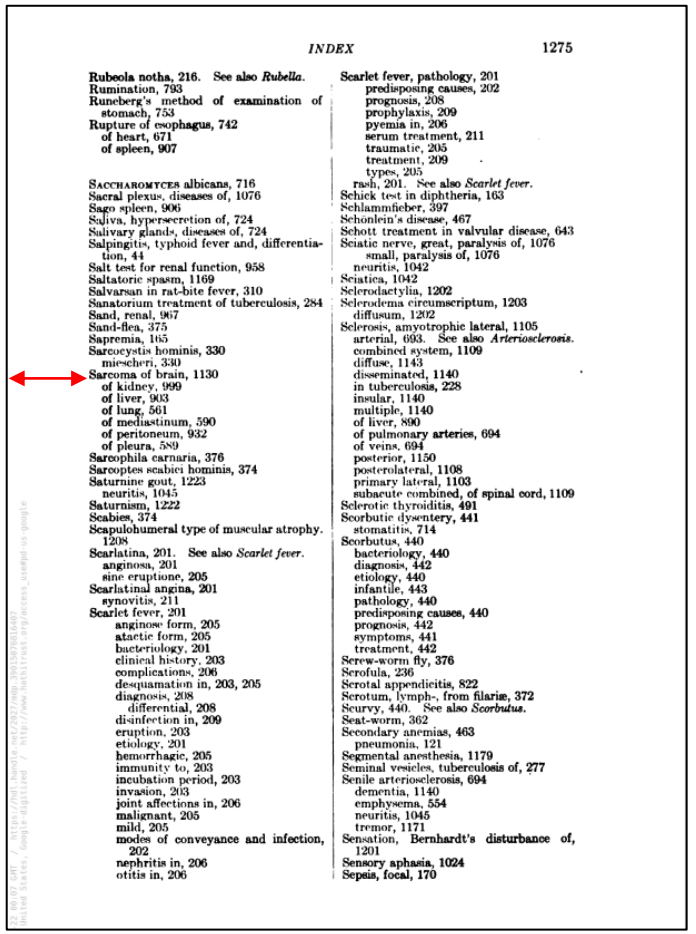
Physical Structure Detection



Technical Challenges

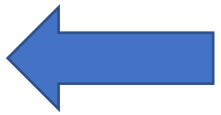


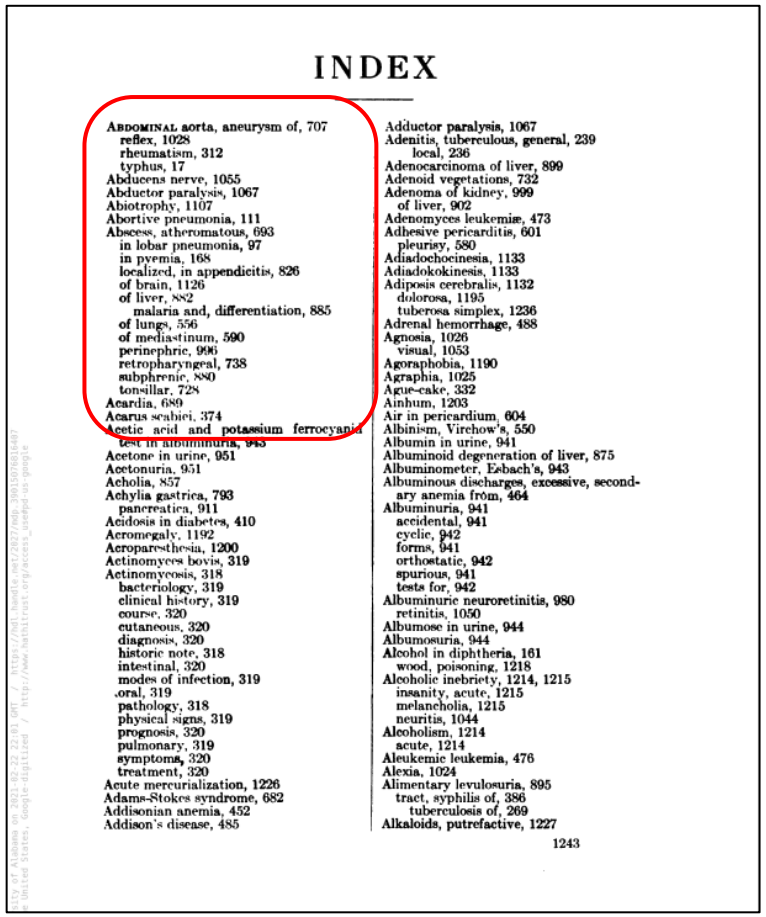
1264 INDEX



Technical Challenges

	X1	Y1	X2	Y2	X3	Y3	X4	Y4	text
2	178	720	740	720	740	757	178	757	ABDOMINAL aorta, aneurysm of, 707
3	211	755	391	754	391	787	211	788	reflex, 1028
4	210	787	477	787	477	821	210	821	rheumatism, 312
5	205	824	379	823	379	855	206	856	typhus, 17
6	178	853	522	853	522	888	178	888	Abducens nerve, 1055
7	179	888	566	887	566	921	179	923	Abductor paralysis, 1067
8	179	922	450	921	450	956	179	957	Abiotrophy, 1107
9	177	954	569	954	569	989	177	989	Abortive pneumonia, 111
10	176	989	606	987	606	1020	176	1022	Abscess, atheromatous, 693
11	205	1020	573	1020	573	1056	205	1056	in lobar pneumonia, 97
12	210	1055	445	1053	445	1087	210	1090	in pvemia, 168
13	210	1085	669	1085	669	1123	210	1123	localized, in appendicitis, 826
14	212	1121	430	1121	430	1154	212	1154	of brain, 1126
15	209	1154	403	1152	403	1185	210	1188	of liver, 882
16	242	1187	749	1185	750	1220	242	1222	malaria and, differentiation, 885
17	207	1220	416	1220	416	1255	207	1255	of lungs, 556
18	209	1254	529	1253	530	1287	209	1288	of mediastinum, 590
19	207	1290	463	1288	464	1319	208	1323	perinephric, 996
20	208	1322	536	1318	536	1354	208	1357	retropharyngeal, 738
21	209	1355	458	1355	458	1386	209	1389	subphrenic, 880
22	208	1388	415	1386	415	1418	208	1419	tonsillar. 728
23	175	1421	373	1421	373	1452	175	1454	Acardia, 689
24	175	1454	475	1453	475	1484	175	1486	Acarus scabiei. 374
25	174	1484	828	1484	828	1520	174	1521	Acetic acid and potassium ferrocyanid



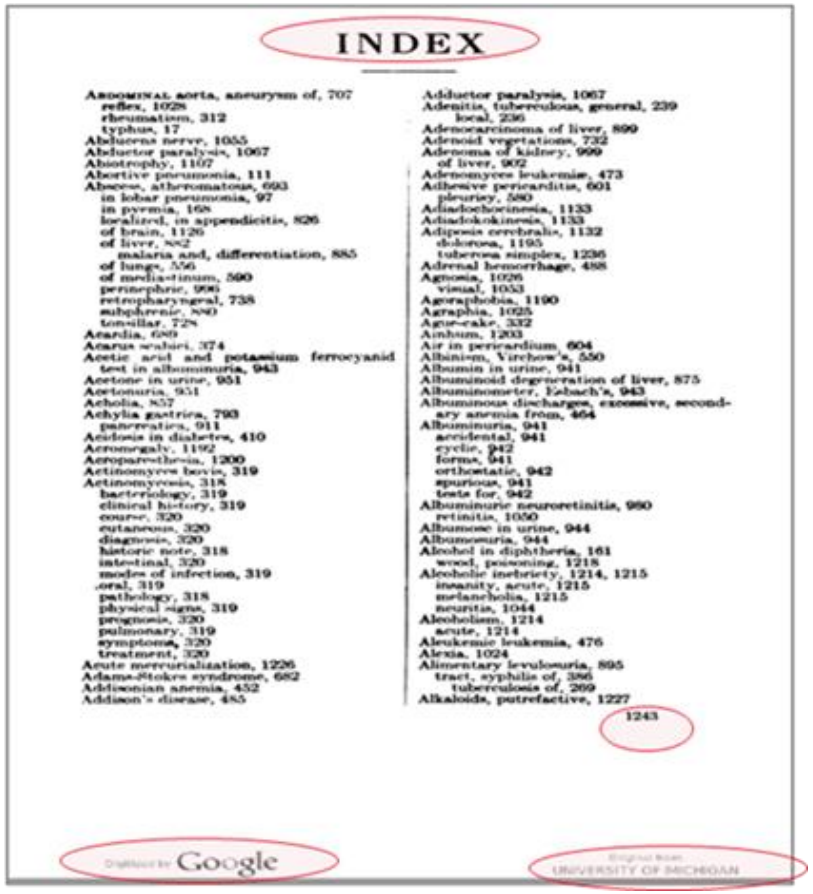


INDEX

ABDOMINAL aorta, aneurysm of, 707
reflex, 1028
rheumatism, 312
typhus, 17
Abducens nerve, 1055
Abductor paralysis, 1067
Abiotrophy, 1107
Abortive pneumonia, 111
Abscess, atheromatous, 693
in lobar pneumonia, 97
in pyemia, 168
localized, in appendicitis, 826
of brain, 1126
of liver, 882
malaria and, differentiation, 885
of lungs, 556
of mediastinum, 590
perinephric, 996
retropharyngeal, 738
subphrenic, 880
tonsillar, 728
Acardia, 689
Acarus scabiei, 374
Acetic acid and potassium ferrocyanid
test in albuminuria, 943
Acetone in urine, 951
Acetonuria, 951
Acholia, 857
Achylia gastrica, 793
pancreatica, 911
Acidosis in diabetes, 410
Acromegaly, 1192
Acroparesthesia, 1200
Actinomyces bovis, 319
Actinomycosis, 318
bacteriology, 319
clinical history, 319
course, 320
cutaneous, 320
diagnosis, 320
historic note, 318
intestinal, 320
modes of infection, 319
oral, 319
pathology, 318
physical signs, 319
prognosis, 320
pulmonary, 319
symptoms, 320
treatment, 320
Acute mercurialization, 1226
Adams-Stokes syndrome, 682
Addisonian anemia, 452
Addison's disease, 485
Adductor paralysis, 1067
Adenitis, tuberculous, general, 239
local, 236
Adenocarcinoma of liver, 899
Adenoid vegetations, 732
Adenoma of kidney, 999
of liver, 902
Adenomyces leukemiæ, 473
Adhesive pericarditis, 601
pleurisy, 580
Adiadochocinesis, 1133
Adiadokokinesis, 1133
Adiposis cerebralis, 1132
dolorosa, 1195
tuberosa simplex, 1236
Adrenal hemorrhage, 488
Agnosia, 1026
visual, 1053
Agoraphobia, 1190
Agraphia, 1025
Ague-cake, 332
Ainhum, 1203
Air in pericardium, 604
Albinism, Virchow's, 550
Albumin in urine, 941
Albuminoid degeneration of liver, 875
Albuminometer, Esbach's, 943
Albuminous discharges, excessive, secondary anemia from, 464
Albuminuria, 941
accidental, 941
cyclic, 942
forms, 941
orthostatic, 942
spurious, 941
tests for, 942
Albuminuric neuroretinitis, 980
retinitis, 1050
Albumose in urine, 944
Albumosuria, 944
Alcohol in diphtheria, 161
wood, poisoning, 1218
Alcoholic inebriety, 1214, 1215
insanity, acute, 1215
melancholia, 1215
neuritis, 1044
Alcoholism, 1214
acute, 1214
Aleukemic leukemia, 476
Alexia, 1024
Alimentary levulosuria, 895
tract, syphilis of, 386
tuberculosis of, 269
Alkaloids, putrefactive, 1227

1243

Noise Elimination



INDEX

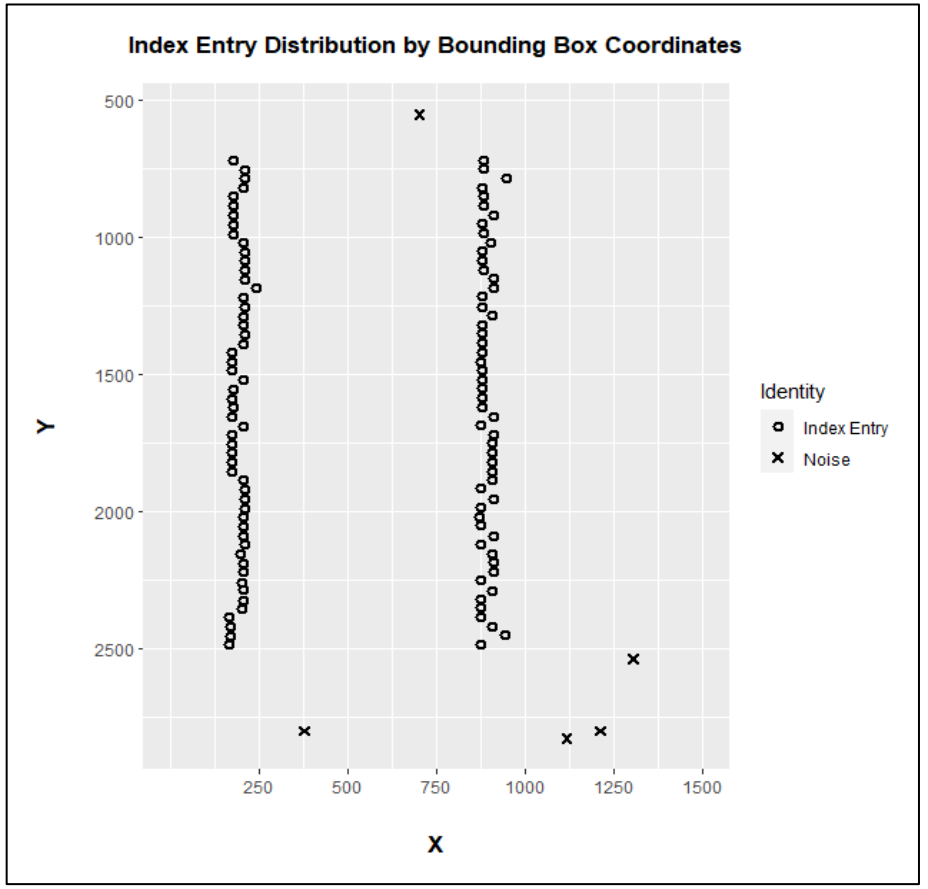
Abdominal aorta, aneurysm of, 707
  reflex, 1028
  rheumatism, 312
  typhus, 17
Abducens nerve, 1055
Abductor paralysis, 1067
Abiotrophy, 1107
Abortive pneumonia, 111
Abscess, atheromatous, 693
  in lobar pneumonia, 97
  in pyemia, 168
  localized, in appendicitis, 826
  of brain, 1126
  of liver, 862
    malaria and, differentiation, 885
  of lungs, 556
  of mediastinum, 560
  perinephric, 996
  retropharyngeal, 738
  subphrenic, 890
  tonsillar, 728
Acardia, 680
Acarus scabiei, 374
Acetic acid and potassium ferrocyanid test in albuminuria, 943
Acetone in urine, 951
Acetonuria, 951
Acholia, 857
Achylia gastrica, 793
  pancreatica, 911
Acidosis in diabetes, 410
Acromegaly, 1192
Acroparesthesia, 1200
Actinomyces bovis, 319
Actinomycosis, 318
  bacteriology, 319
  clinical history, 319
  course, 320
  cutaneous, 320
  diagnosis, 320
  historic note, 318
  intestinal, 320
  modes of infection, 319
  oral, 319
  pathology, 318
  physical signs, 319
  prognosis, 320
  pulmonary, 319
  symptoms, 320
  treatment, 320
Acute mercurialization, 1226
Adams-Stokes syndrome, 682
Addisonian anemia, 452
Addison's disease, 485
Adductor paralysis, 1067
Adenitis, tuberculous, general, 239
  local, 236
Adenocarcinoma of liver, 899
Adenoid vegetations, 732
Adenoma of kidney, 999
  of liver, 902
Adenomyces leukemia, 473
Adhesive pericarditis, 601
  pleurisy, 580
Adiadochocinesis, 1133
Adiadokokinesis, 1133
Adiposis cerebralis, 1132
  dolorosa, 1195
  tuberosa simplex, 1236
Adrenal hemorrhage, 488
Agnosia, 1026
  visual, 1053
Agoraphobia, 1190
Agraphia, 1025
Ague-cake, 332
Ainhum, 1203
Air in pericardium, 604
Albinism, Virchow's, 550
Albumin in urine, 941
Albuminoid degeneration of liver, 875
Albuminometer, Esbach's, 943
Albuminous discharges, excessive, secondary anemia from, 464
Albuminuria, 941
  accidental, 941
  cyclic, 942
  forms, 941
  orthostatic, 942
  spurious, 941
  tests for, 942
Albuminuric neuroretinitis, 990
  retinitis, 1050
Albumose in urine, 944
Albumosuria, 944
Alcohol in diphtheria, 161
  wood, poisoning, 1218
Alcoholic inebriety, 1214, 1215
  insanity, acute, 1215
  melancholia, 1215
  neuritis, 1044
Alcoholism, 1214
  acute, 1214
Aleukemic leukemia, 476
Alexia, 1024
Alimentary levulosuria, 895
  tract, syphilis of, 396
  tuberculosis of, 269
Alkaloids, putrefactive, 1227

1243

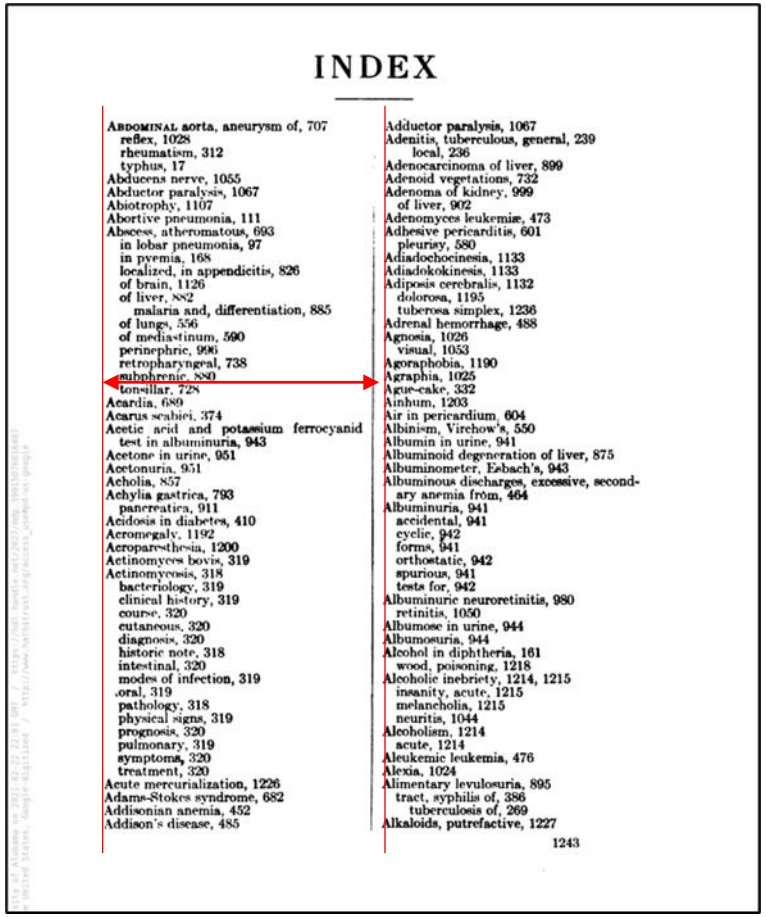
Digitized by Google

Original from UNIVERSITY OF MICHIGAN

DBSCAN Algorithm



Column Segmentation

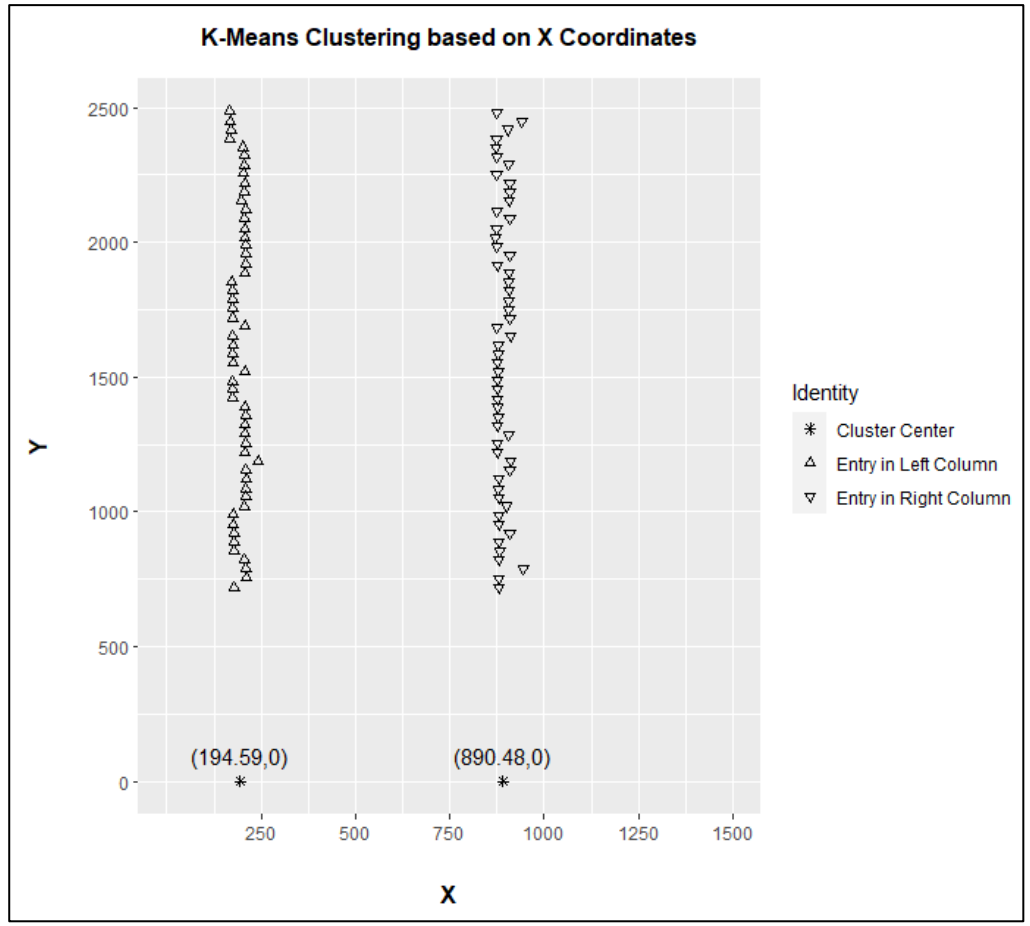


INDEX

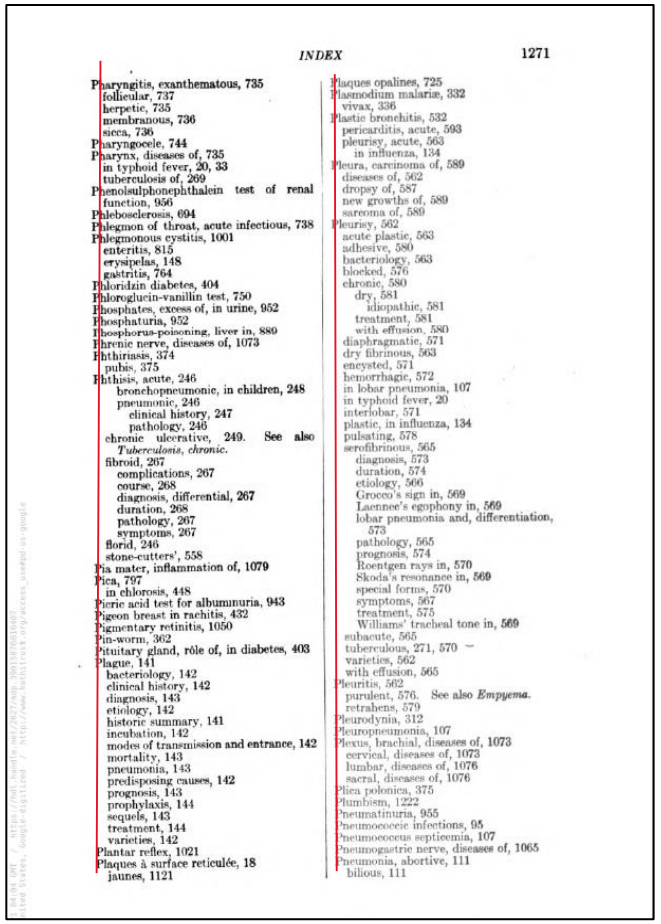
Abdominal aorta, aneurysm of, 707
reflex, 1028
rheumatism, 312
typhus, 17
Abducens nerve, 1055
Abductor paralysis, 1067
Abiotrophy, 1107
Abortive pneumonia, 111
Abscess, atheromatous, 693
in lobar pneumonia, 97
in pyemia, 168
localized, in appendicitis, 826
of brain, 1126
of liver, 882
malaria and, differentiation, 885
of lungs, 556
of mediastinum, 590
perinephric, 996
retropharyngeal, 738
subphrenic, 880
tonsillar, 728
Acardia, 689
Acarus scabiei, 374
Acetic acid and potassium ferrocyanid test in albuminuria, 943
Acetone in urine, 951
Acetonuria, 951
Acholia, 857
Achylia gastrica, 793
pancreatica, 911
Acidosis in diabetes, 410
Acromegaly, 1192
Acroparesthesia, 1200
Actinomyces bovis, 319
Actinomycosis, 318
bacteriology, 319
clinical history, 319
course, 320
cutaneous, 320
diagnosis, 320
historic note, 318
intestinal, 320
modes of infection, 319
oral, 319
pathology, 318
physical signs, 319
prognosis, 320
pulmonary, 319
symptoms, 320
treatment, 320
Acute mercurialization, 1226
Adams-Stokes syndrome, 682
Addisonian anemia, 452
Addison's disease, 485
Adductor paralysis, 1067
Adenitis, tuberculous, general, 239
local, 236
Adenocarcinoma of liver, 899
Adenoid vegetations, 732
Adenoma of kidney, 999
of liver, 902
Adenomyces leukemiæ, 473
Adhesive pericarditis, 601
pleurisy, 580
Adiadochocinesis, 1133
Adiadokokinesis, 1133
Adiposis cerebralis, 1132
dolorosa, 1195
tuberosa simplex, 1236
Adrenal hemorrhage, 488
Agnosia, 1026
visual, 1053
Agoraphobia, 1190
Agraphia, 1025
Ague-cake, 332
Ainhum, 1203
Air in pericardium, 604
Albinism, Virchow's, 550
Albumin in urine, 941
Albuminoid degeneration of liver, 875
Albuminometer, Esbach's, 943
Albuminous discharges, excessive, secondary anemia from, 464
Albuminuria, 941
accidental, 941
cyclic, 942
forms, 941
orthostatic, 942
spurious, 941
tests for, 942
Albuminuric neuroretinitis, 980
retinitis, 1050
Albumose in urine, 944
Albumosuria, 944
Alcohol in diphtheria, 161
wood, poisoning, 1218
Alcoholic inebriety, 1214, 1215
insanity, acute, 1215
melancholia, 1215
neuritis, 1044
Alcoholism, 1214
acute, 1214
Aleukemic leukemia, 476
Alexia, 1024
Alimentary levulosuria, 895
tract, syphilis of, 386
tuberculosis of, 269
Alkaloids, putrefactive, 1227

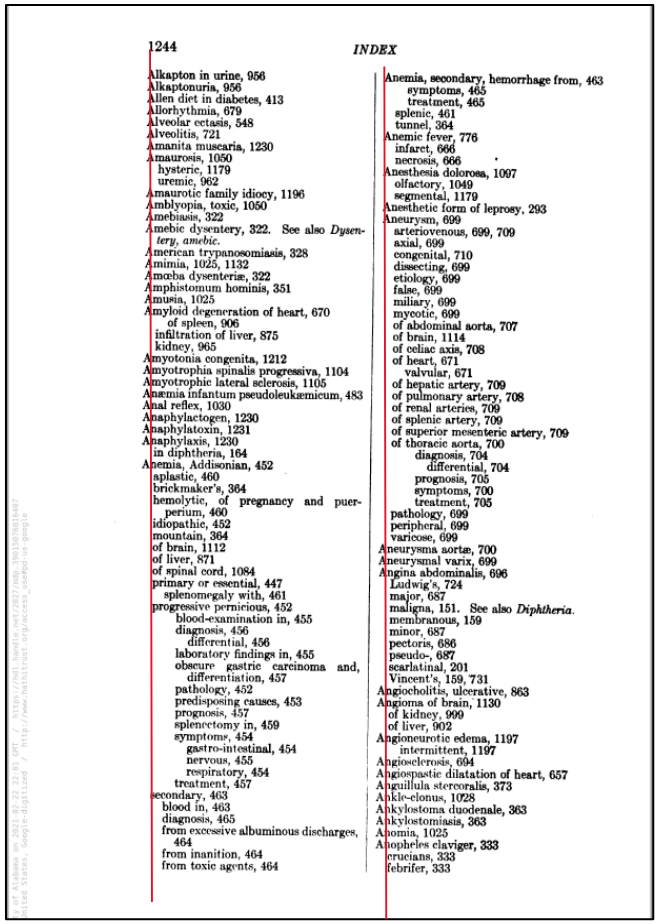
1243

K-Means Algorithm



Skew Correction

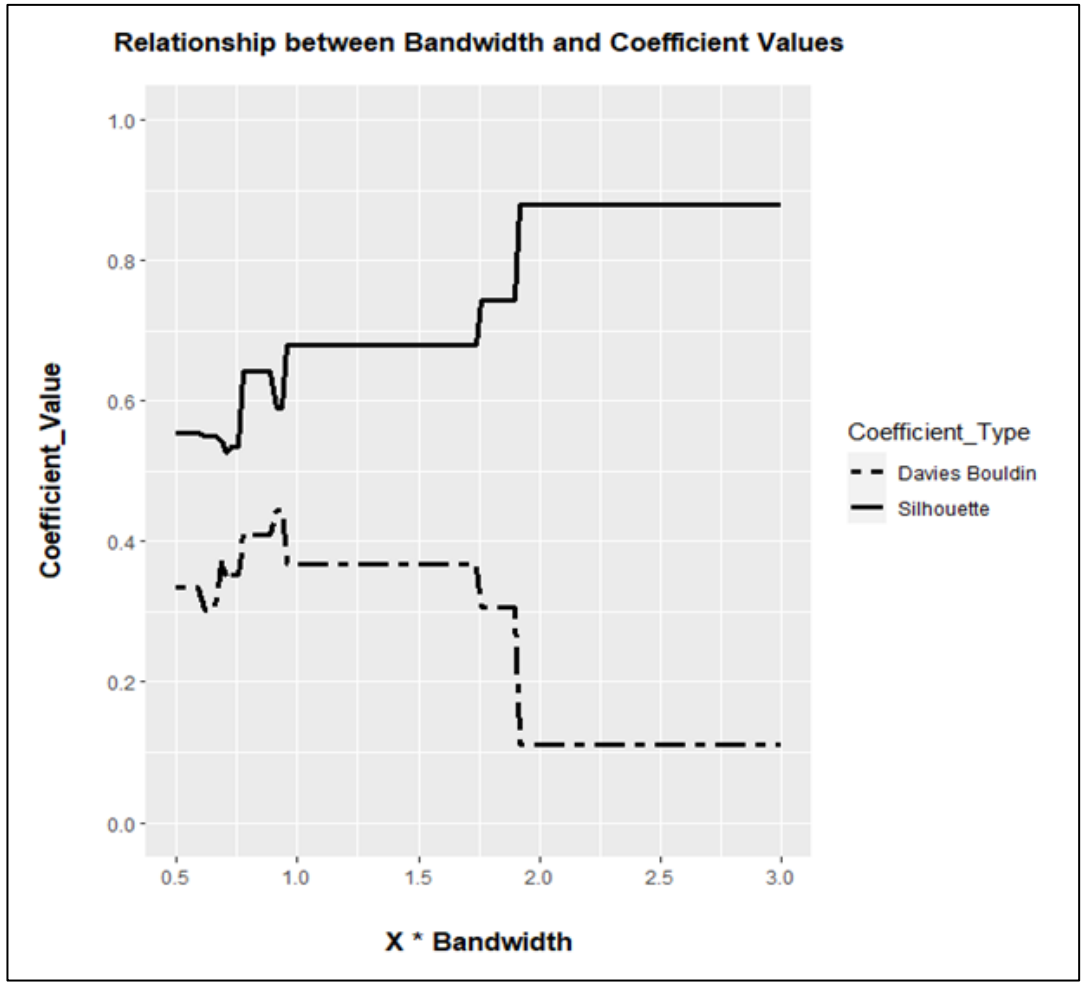




$$\tan(\theta^\circ) = 1/N \cdot \sum (Y_R - Y_L)/(X_R - X_L) \qquad (1)$$

$$X_i' = X_i + \tan(\theta^\circ) \cdot (Y_i - Y_1) \qquad (2)$$

Entry Classification

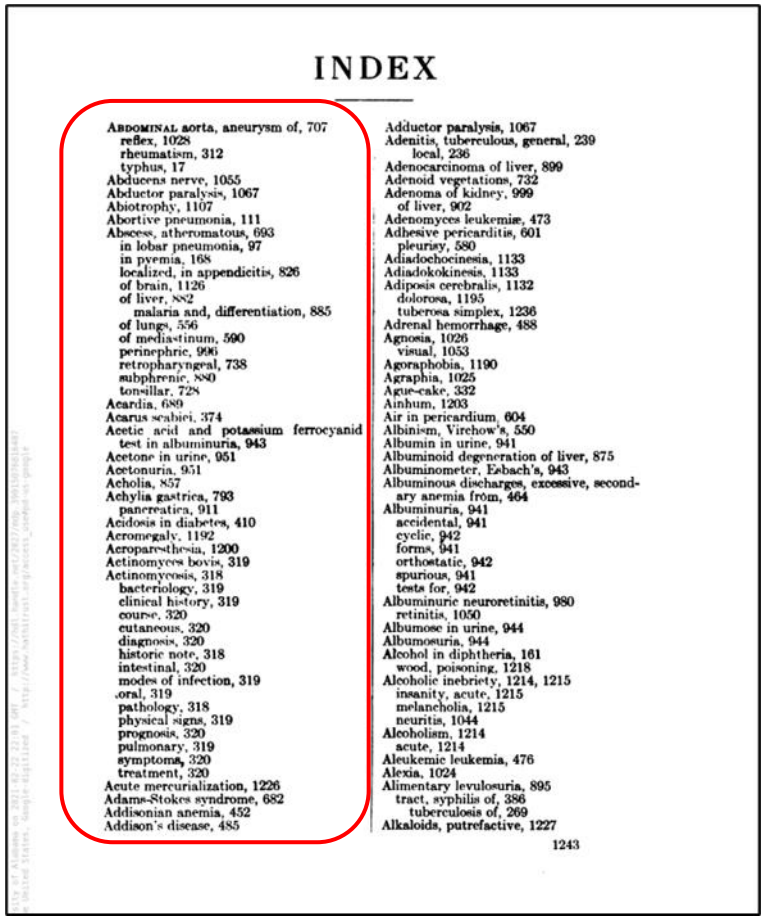


Mean Shift Bandwidth Estimation

Mean Shift clustering performance evaluation:

- Silhouette Coefficient
- Davies-Bouldin Index

Recovering Typographical Layout Logic with

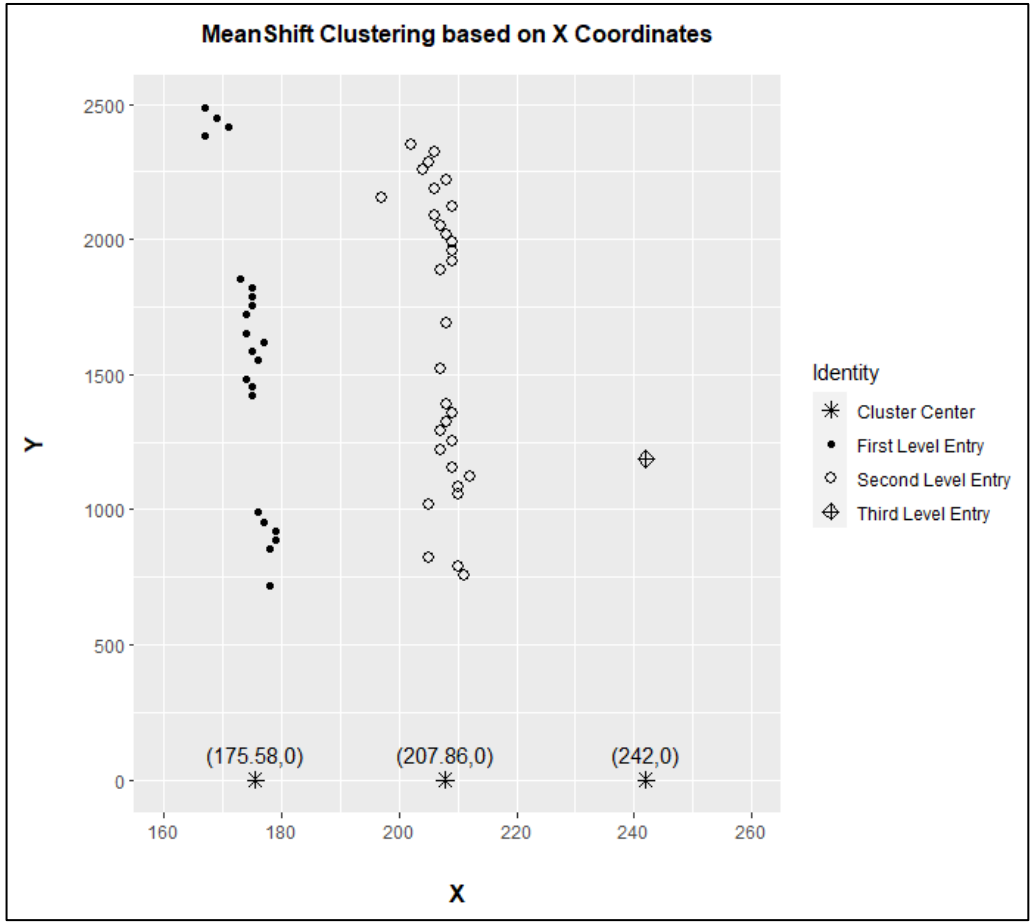


INDEX

Abdominal aorta, aneurysm of, 707
reflex, 1028
rheumatism, 312
typhus, 17
Abducens nerve, 1055
Abductor paralysis, 1067
Abiotrophy, 1107
Abortive pneumonia, 111
Abscess, atheromatous, 693
in lobar pneumonia, 97
in pyemia, 168
localized, in appendicitis, 826
of brain, 1126
of liver, 882
malaria and, differentiation, 885
of lungs, 556
of mediastinum, 590
perinephric, 996
retropharyngeal, 738
subphrenic, 880
tonsillar, 728
Acardia, 689
Acarus scabiei, 374
Acetic acid and potassium ferrocyanid test in albuminuria, 943
Acetone in urine, 951
Acetonuria, 951
Acholia, 857
Achylia gastrica, 793
pancreatica, 911
Acidosis in diabetes, 410
Acromegaly, 1192
Acroparesthesia, 1200
Actinomyces bovis, 319
Actinomycosis, 318
bacteriology, 319
clinical history, 319
course, 320
cutaneous, 320
diagnosis, 320
historic note, 318
intestinal, 320
modes of infection, 319
oral, 319
pathology, 318
physical signs, 319
prognosis, 320
pulmonary, 319
symptoms, 320
treatment, 320
Acute mercurialization, 1226
Adams-Stokes syndrome, 682
Addisonian anemia, 452
Addison's disease, 485
Adductor paralysis, 1067
Adenitis, tuberculous, general, 239
local, 236
Adenocarcinoma of liver, 899
Adenoid vegetations, 732
Adenoma of kidney, 999
of liver, 902
Adenomyces leukemiæ, 473
Adhesive pericarditis, 601
pleurisy, 580
Adiadochocinesis, 1133
Adiadokokinesis, 1133
Adiposis cerebralis, 1132
dolorosa, 1195
tuberosa simplex, 1236
Adrenal hemorrhage, 488
Agnosia, 1026
visual, 1053
Agoraphobia, 1190
Agraphia, 1025
Ague-cake, 332
Ainhum, 1203
Air in pericardium, 604
Albinism, Virchow's, 550
Albumin in urine, 941
Albuminoid degeneration of liver, 875
Albuminometer, Esbach's, 943
Albuminous discharges, excessive, secondary anemia from, 464
Albuminuria, 941
accidental, 941
cyclic, 942
forms, 941
orthostatic, 942
spurious, 941
tests for, 942
Albuminuric neuroretinitis, 980
retinitis, 1050
Albumose in urine, 944
Albumosuria, 944
Alcohol in diphtheria, 161
wood, poisoning, 1218
Alcoholic inebriety, 1214, 1215
insanity, acute, 1215
melancholia, 1215
neuritis, 1044
Alcoholism, 1214
acute, 1214
Aleukemic leukemia, 476
Alexia, 1024
Alimentary levulosuria, 895
tract, syphilis of, 386
tuberculosis of, 269
Alkaloids, putrefactive, 1227

1243

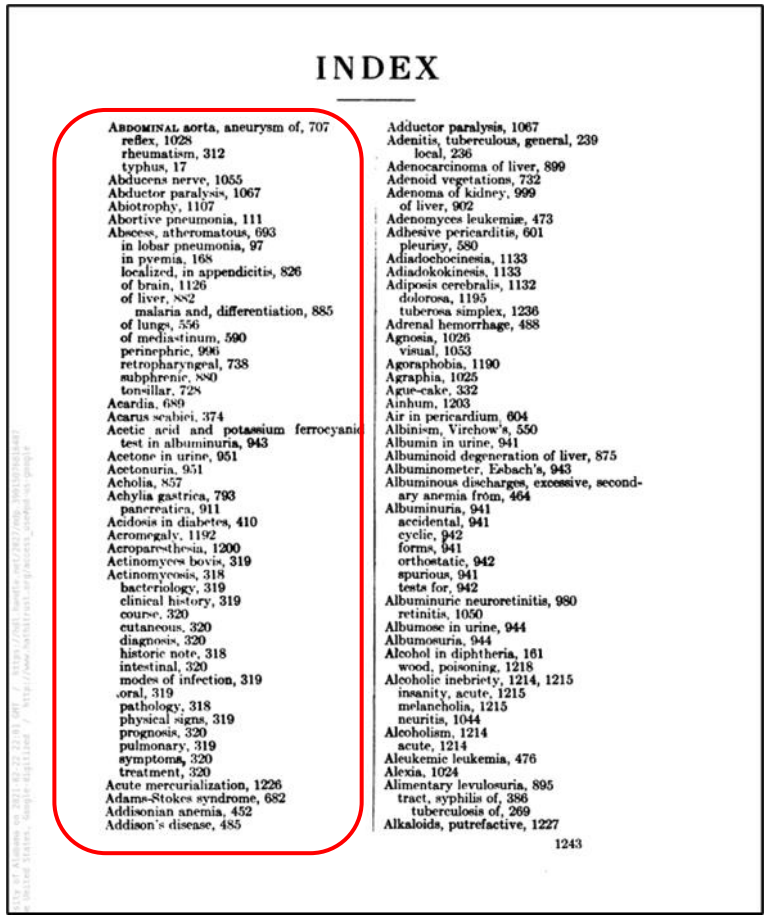
Mean Shift Algorithm



Evaluation of Pipeline Effectiveness: Accuracy Testing

	Entry_Text	Entry_Level
1	ABDOMINAL aorta, aneurysm of, 707	1
2	reflex, 1028	2
3	rheumatism, 312	2
4	typhus, 17	2
5	Abducens nerve, 1055	1
6	Abductor paralysis, 1067	1
7	Abiotrophy, 1107	1
8	Abortive pneumonia, 111	1
9	Abscess, atheromatous, 693	1
10	in lobar pneumonia, 97	2
11	in pvemia, 168	2
12	localized, in appendicitis, 826	2
13	of brain, 1126	2
14	of liver, 882	2
15	malaria and, differentiation, 885	3
16	of lungs, 556	2
17	of mediastinum, 590	2
18	perinephric, 996	2
19	retropharyngeal, 738	2
20	subphrenic, 880	2
21	tonsillar. 728	2
22	Acardia, 689	1
23	Acarus scabiei. 374	1
24	Acetic acid and potassium ferrocyanid	1
25	test in albuminuria, 943	2

Accuracy Testing



INDEX

ABDOMINAL aorta, aneurysm of, 707
reflex, 1028
rheumatism, 312
typhus, 17
Abducens nerve, 1055
Abductor paralysis, 1067
Abiotrophy, 1107
Abortive pneumonia, 111
Abscess, atheromatous, 693
in lobar pneumonia, 97
in pyemia, 168
localized, in appendicitis, 826
of brain, 1126
of liver, 882
malaria and, differentiation, 885
of lungs, 556
of mediastinum, 590
perinephric, 996
retropharyngeal, 738
subphrenic, 880
tonsillar, 728
Acardia, 689
Acarus scabiei, 374
Acetic acid and potassium ferrocyanid
test in albuminuria, 943
Acetone in urine, 951
Acetonuria, 951
Acholia, 857
Achylia gastrica, 793
pancreatica, 911
Acidosis in diabetes, 410
Acromegaly, 1192
Acroparesthesia, 1200
Actinomyces bovis, 319
Actinomycosis, 318
bacteriology, 319
clinical history, 319
course, 320
cutaneous, 320
diagnosis, 320
historic note, 318
intestinal, 320
modes of infection, 319
oral, 319
pathology, 318
physical signs, 319
prognosis, 320
pulmonary, 319
symptoms, 320
treatment, 320
Acute mercurialization, 1226
Adams-Stokes syndrome, 682
Addisonian anemia, 452
Addison's disease, 485
Adductor paralysis, 1067
Adenitis, tuberculous, general, 239
local, 236
Adenocarcinoma of liver, 899
Adenoid vegetations, 732
Adenoma of kidney, 999
of liver, 902
Adenomyces leukemiæ, 473
Adhesive pericarditis, 601
pleurisy, 580
Adiadochocinesia, 1133
Adiadokokinesis, 1133
Adiposis cerebralis, 1132
dolorosa, 1195
tuberosa simplex, 1236
Adrenal hemorrhage, 488
Agnosia, 1026
visual, 1053
Agoraphobia, 1190
Agraphia, 1025
Ague-cake, 332
Ainhum, 1203
Air in pericardium, 604
Albinism, Virchow's, 550
Albumin in urine, 941
Albuminoid degeneration of liver, 875
Albuminometer, Esbach's, 943
Albuminous discharges, excessive, secondary anemia from, 464
Albuminuria, 941
accidental, 941
cyclic, 942
forms, 941
orthostatic, 942
spurious, 941
tests for, 942
Albuminuric neuroretinitis, 980
retinitis, 1050
Albumose in urine, 944
Albumosuria, 944
Alcohol in diphtheria, 161
wood, poisoning, 1218
Alcoholic inebriety, 1214, 1215
insanity, acute, 1215
melancholia, 1215
neuritis, 1044
Alcoholism, 1214
acute, 1214
Aleukemic leukemia, 476
Alexia, 1024
Alimentary levulosuria, 895
tract, syphilis of, 386
tuberculosis of, 269
Alkaloids, putrefactive, 1227

1243

Evaluation of Pipeline Effectiveness: ML Procedures

Procedure	Algorithm	Accuracy
Noise Elimination	DBSCAN	95.4 %
Column Segmentation	K-Means	98.2 %
Entry Classification	Mean Shift	93.8 %

Recovered Typographical Layout Logic to Graph (1)

Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c) (Q11494)

From Wikibase.slis.ua.edu

From: Modern medicine : its theory and practice, in original contributions by American and foreign authors (1907-1910.). Volume 1.

⟨wikib

In more languages Configure

Language	Label	Description	Also known as
English	Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c)	From: Modern medicine : its theory and practice, in original contributions by American and foreign authors (1907-1910.). Volume 1.	

Statements

instance of	Book index main heading 0 references
has part	==> demodectic, 630 (Index entry from HathiTrust item t54f2j58c) 0 references
	==>sarcoptic, 627 (Index entry from HathiTrust item t54f2j58c) 0 references

Recovered Typographical Layout Logic to Graph (2)

==> demodectic, 630 (Index entry from HathiTrust item t54f2j58c) (Q11495)

From Wikibase.slis.ua.edu

From: Modern medicine : its theory and practice, in original contributions by American and foreign authors (1907-1910.). Volume 1.

In more languages Configure

Language	Label	Description	Also known as
English	==> demodectic, 630 (Index entry from HathiTrust item t54f2j58c)	From: Modern medicine : its theory and practice, in original contributions by American and foreign authors (1907-1910.). Volume 1.	

Statements

instance of	Book index subheading 0 references
part of	Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c) 0 references
has part	====>parasite of, 630 (Index entry from HathiTrust item t54f2j58c) 0 references
	====>symptoms of, 630 (Index entry from HathiTrust item t54f2j58c) 0 references

Displaying Typographical Logic Layout: SPARQL Query

- The transitive operator (*) allows for queries that follow our partOf (P22) property path, which captures the multiple levels of index subentries. under the Acariasis (Q11494) main entry in book index.
- Transitive SPARQL query example (https://tinyurl.com/ydv6rq2y):

```
# Book index entry for "Acariasis": Example query with transitive operator demonstrating property path search
SELECT ?IndexEntryLabel ?URL
WHERE
{
  wd:Q11494 wdt:P22* ?IndexEntry.
  ?IndexEntry wdt:P181 ?URL.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

Displaying Typographical Logic Layout: SPARQL Results

Arrows indicate restored indentations. Wikibase SPARQL query results exportable as HTML for overlay index.

IndexEntryLabel	URL
Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=573>
Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=668>
Acariasis, 533, 626, 631 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=673>
==> demodectic, 630 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=672>
====>parasite of, 630 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=672>
====>symptoms of, 630 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=672>
==>sarcoptic, 627 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=669>
====>diagnosis of, 629 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=671>
====>frequency of, 628 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=670>
====>parasites of, 627 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=669>
====>symptoms of, 628 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=670>
====>treatment of, 629 (Index entry from HathiTrust item t54f2j58c)	<https://babel.hathitrust.org/cgi/pt?id=nnc2.ark:/13960/t54f2j58c&view=1up&seq=671>

Conclusion and Future Work

- Conclusion”
  - Pipeline development: From theory to optimizing for scale up
  - Evaluation

- Future work:
  - Layout variations
  - Overlay index interface work in the context of a larger expression of navigation paratext theory and philology graphs

References

- Birke, Dorothee and Christ, Birte. 2013. Paratext and Digitized Narrative: Mapping the Field. Narrative 21(1), 65-87.
- Genette, Gerard. 1997. Paratexts: thresholds of interpretation. Cambridge University Press.
- Klement, Susan. 2002. Open-system versus closed-system indexing. The Indexer, 23(1), 23-31.
- Mayer-Schönberger, Viktor and Cukier, Kenneth. 2013. Chapter 5. Datafication. In Big Data : A Revolution That Will Transform How We Live, Work, and Think (pp. 73-97). Houghton, Mifflin, Harcourt.