

Data Mining, Visualizing, and Analyzing Faculty Thematic Relationships for Research Support and Collection Analysis

ASSOCIATION OF COLLEGE & RESEARCH LIBRARIES CONFERENCE 2017

Wenli Gao, Communication, Sociology, Anthropology Librarian, University of Houston

Loretta Wallace, Business Librarian, University of Texas at Austin



Freeimages.com/israylov

Why are we doing this?

Process

Scopus



topic-modeling-tool



Data
Collection



















Data
Processing

Word
Frequencies

Topic
Modeling

Data
Visualization

Results

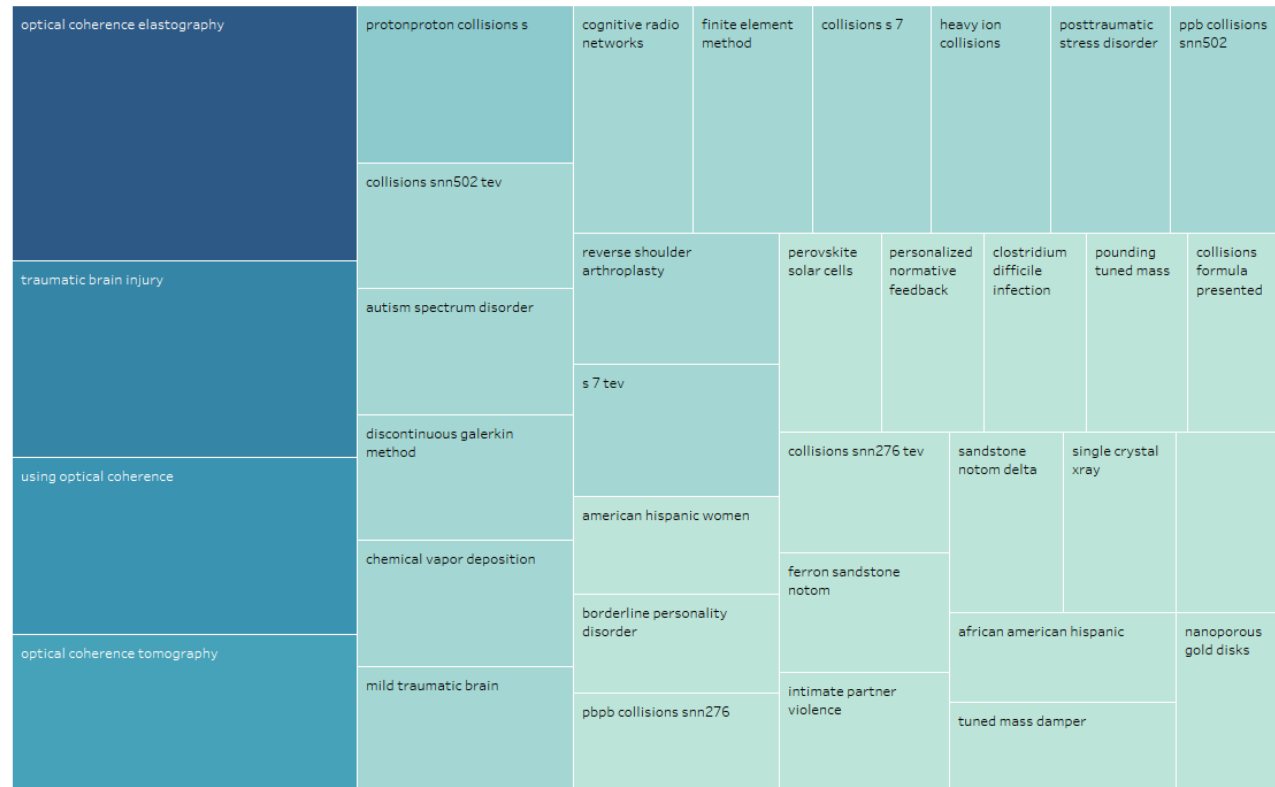
 06title.csv
 06title.csv_1words.csv
 06title.csv_2words.csv
 06title.csv_3words.csv
 06title.csv_4words.csv
 06topic.csv
 07title.csv
 07title.csv_1words.csv
 07title.csv_2words.csv
 07title.csv_3words.csv
 07title.csv_4words.csv
 07topic.csv
 08titles.csv
 08titles.csv_1words.csv
 08titles.csv_2words.csv
 08titles.csv_3words.csv
 08titles.csv_4words.csv
 08topic.csv

phrase	count
('sum', 'frequency')	10
('frequency', 'generation')	10
('magnetic', 'recording')	8
('carbon', 'nanotubes')	8
('vibration', 'control')	7
('wireless', 'networks')	6
('english', 'language')	5
('shape', 'memory')	5
('scattering', 'series')	5
('drug', 'diffusion')	5
('neonatal', 'seizures')	5
('vitro', 'experimental')	4
('risk', 'factors')	4
('experimental', 'conditions')	4
('dy', 'ho')	4
('conditions', 'drug')	4
('memory', 'alloys')	4
('surface', 'plasmon')	4
('videotaped', 'neonatal')	4
('oxide', 'films')	4
('university', 'houston')	4
('carbon', 'nanofiber')	4
('diffusion', 'cornea')	4
('learners', 'risk')	4
('case', 'study')	4
('neural', 'networks')	4
('controller', 'design')	4
('finite', 'element')	4
('heat', 'conduction')	4
('sensor', 'networks')	4

Phrase	Count	Year	Number of Words
analysis	54	2006	1
model	37	2006	1
systems	36	2006	1
design	33	2006	1
effects	31	2006	1
control	31	2006	1
properties	30	2006	1
effect	29	2006	1
networks	28	2006	1
surface	26	2006	1
new	24	2006	1
magnetic	23	2006	1
films	22	2006	1
data	22	2006	1
synthesis	21	2006	1
modeling	21	2006	1
characteri	21	2006	1
approach	21	2006	1
structural	20	2006	1
performan	20	2006	1
seismic	19	2006	1
field	19	2006	1
applicatio	19	2006	1
water	18	2006	1
risk	18	2006	1
dynamics	18	2006	1
detection	18	2006	1
structures	17	2006	1

Visualization

Faculty Publications Word Frequency



Number of Words

- ☐ (All)
- ☐ 1
- ☒ 2
- ☐ 3
- ☐ 4

Year

- ☐ (All)
- ☐ 2006
- ☐ 2007
- ☐ 2008
- ☐ 2009
- ☐ 2010
- ☐ 2011
- ☐ 2012
- ☐ 2013
- ☐ 2014
- ☒ 2015

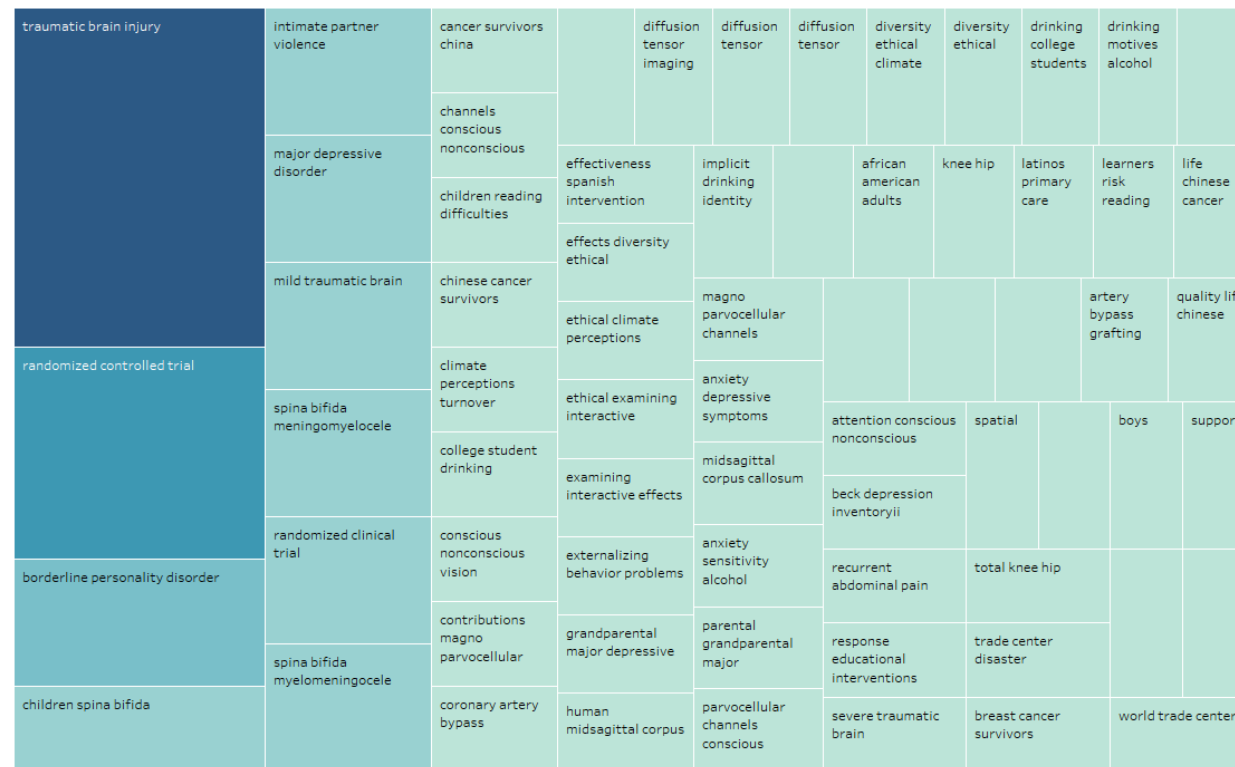
Count



59

Visualization

Faculty Publications Word Frequency by Department



Number of Words

- ☐ (All)
☐ 1
☐ 2
☒ 3
☐ 4

Year

- ☒ (All)
☒ 2006
☒ 2007
☒ 2008
☒ 2009
☒ 2010
☒ 2011
☒ 2012
☒ 2013
☒ 2014
☒ 2015

Department

- ☐ (All)
☐ Chemistry
☐ Computer Science
☒ Psychology

What are the benefits?





What were the limitations?

What are the next steps?



Questions

Wenli Gao, Social Science Liaison, University of Houston
wgao5@uh.edu

Loretta Wallace, Business Liaison, University of Texas at Austin
l.wallace@austin.utexas.edu