True North Science
Bootcamp 2018

Holly Hendrigan
@hollyhen
24 May 2018

Digital humanities and STEM librarianship,

or why I stopped rolling my eyes at word clouds.

Overview

- Background
- Methodological failures
  - Google Alerts
  - Mallet
- Methodological success
  - Web of Science
  - Voyant Tools
- Key findings / results
  - Text mining theory
- Reservations

Background

SFU SIMON FRASER UNIVERSITY
LIBRARY

Liaison Program Design Team (LPDT)
Report and Recommendations July 2015

Design Team

Megan Crouch
Elaine Fairey
Patty Gallilee
Karen Marotz
Shane Plante
Hope Power
Sylvia Roberts
Nina Saklikar
Nicole White
Baharak Yousefi

Proposed Subject Grouping C

12.5 FTE
Each grouping below 1.0 FTE unless otherwise indicated.

BPK
Biology
MBB
Health

Business/Economics

Business
Urban Studies
Public Policy

Contemporary Arts
Communications

Earth Sciences
Statistics
Physics
Math
Chemistry

Education

English
World Literature
French
Linguistics
Humanities

Environment*
REM
Archaeology
First Nations Department

Gerontology
Sociology/Anthropology
GSWS

International Studies
Political Science
History
Liberal Studies

Mechatronics
Engineering
Computing Science

Psychology
Criminology
Philosophy

SIAT
Publishing

* Grouping considered 0.5 FTE paired with the Indigenous Initiatives role.

STEM librarian onboarding

- Maness, J (2016); Tobin Cataldo et al (2006)
  - Network with STEM librarian colleagues
  - Read websites
  - Conduct selective faculty literature searches
  - Build relationships

Faculty Website

WHAT IS MECHATRONIC SYSTEMS ENGINEERING?

Mechatronics is a dynamic, multidisciplinary subject combining three engineering fields: mechanical, electrical and software engineering. This highly integrated approach creates smart, inventive and evermore efficient solutions for a wide range of high-tech engineering problems.

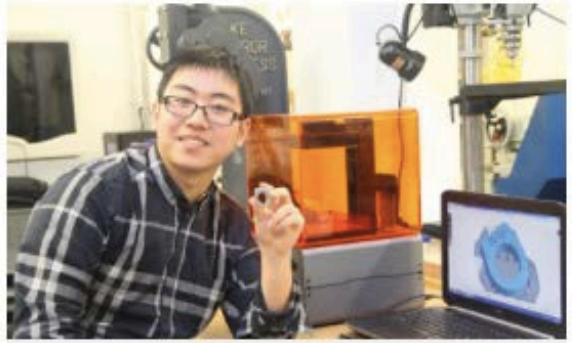
Research interests of Professor X:

Research Interests

- Mechatronics,
- Biomechatronics,
- Biomedical technologies,
- Wearable technologies,
- Biorobotics

Broad categories of research areas

BIO MECHATRONICS



COMPUTATIONAL DESIGN & ENGINEERING

ENERGY SYSTEMS

INDUSTRIAL & POWER ELECTRONICS

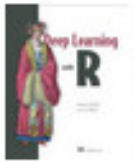
INTELLIGENT SYSTEMS & CONTROL

MICRO & NANO TECHNOLOGIES

Varadi, Peter F. - The sun is rising in Africa and the Middle East : on the road to a solar energy future
ISBN/EAN: 9789814774895 [9814774898] Order Info
(Paperback 69.95 USD - 14% / 77.62 CAD estimated net price) (Customer Demand: 1)
Usually ships in 4+ weeks *Approval Treatment* *Alternate eBooks*
Initial Selector: **H. Hendrigan**

Descriptive analysis in sensory evaluation
ISBN/EAN: 9780470671399 [0470671394] Order Info
(Cloth/HB 200.00 USD - 14% / 221.95 CAD estimated net price) (Customer Demand)
In stock *Approval Treatment* *Alternate eBooks*
Initial Selector: **H. Hendrigan**

Main group strategies towards functional organic materials
ISBN/EAN: 9781119235972 [1119235979] *In Slip Notifications* Order Info
(Cloth/HB 185.00 USD - 14% / 205.30 CAD estimated net price) (Customer Demand: 2)
Usually ships in 1 – 2 weeks* *Approval Treatment* *Alternate eBooks*
Initial Selector: **H. Hendrigan**



Chollet, François. - Deep learning with R
ISBN/EAN: 9781617295546 [161729554X] Order Info
(Paperback 49.99 USD - 14% / 55.48 CAD estimated net price) (Customer Demand: 5)
In stock *Approval Treatment*
Initial Selector: **H. Hendrigan**

Skabara, Peter - Nanostructured materials for type III photovoltaics. edited by Peter Skabara, Mohammad Azad Malik
ISBN/EAN: 9781782624585 [1782624589] *In Your Slip Notifications* Order Info
(Cloth/HB 300.00 CAD - 14% / 258.00 CAD estimated net price) (Customer Demand: 2)
In stock *Approval Treatment* *Alternate eBooks*
Initial Selector: **H. Hendrigan**

Muret, Pierre - Fundamentals of electronics 2 : Continuous-time signals and systems
ISBN/EAN: 9781786301826 [1786301822] *In Your Slip Notifications* Order Info
(Cloth/HB 130.00 USD - 14% / 144.27 CAD estimated net price) (Customer Demand: 1)
Usually ships in 4+ weeks *Approval Treatment* *Alternate eBooks* *Alternate Print*
Initial Selector: **H. Hendrigan**

Addressing the Knowledge Gap 1: Google Scholar

Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -
Scholar Alert - ["school of engineering science simon fraser university"] -

Excel Spreadsheet

	A	B	C	E
1	**Researcher**	**Article Title**	**Journal Title**	**Keywords**
2	Hyomin Choi & Ivan V. E	HEVC INTRA FEATURES FOR HUMAN DETECTION	To be presented at IEEE GlobalSIP'17, Montreal, QC, Nov	Human detection, SVM, HEVC, intra coding
3	Udit Pareek & Daniel C.	Monotonic Optimization for Power Assignment in Two-Way Cognitive Radio Networks with Shared-Band Amplify-and-Forward Relays	2017 26th International Conference on Computer Communication and Networks (ICCCN)	Relays, Approximation algorithms, Cognitive radio, Interference, Optimization methods,Cooperative communication
4	Chinthaka Dinesh, Step	INCORPORATING TIME-OF-DAY USAGE PATTERNS INTO NON-INTRUSIVE LOAD MONITORING	Accepted at the 2017 5th IEEE Global Conference on Signal and Information Processing (GlobalSIP).	non-intrusive load monitoring (NILM), time-of-day usage, smart grid, load disaggregation, graph spectral representation
5	Saeed Ranjbar Alvar, Hy	Can you tell a face from a HEVC bitstream?	arXiv.org-Computer Vision and Pattern Recognition (cs.(	face detection, HEVC, deep learning, convolutional neural network
6	Daehan Chung and Bon	Printing-based fabrication method using sacrificial paper substrates for flexible and wearable microfluidic devices	Journal of Micromechanics and Microengineering	printing-based fabrication process, Microfluidic devices, clothing-based wearable microfluidic sensors
7	Jie Liang	Single depth image super-resolution with multiple residual dictionary learning and refinement	2017 IEEE International Conference on Multimedia and E	Image resolution, Dictionaries, Image edge detection, Image reconstruction, Machine learning, Interpolation, Training, Depth image, super-resolution, residual-dictionary learning, shape-adaptive, weighted median filtering
8	Jie Liang	A new combined PSNR for objective video quality assessment	2017 IEEE International Conference on Multimedia and E	Sensitivity, Video recording, Quality assessment, Encoding, Encoding, Encoding, Visualization, Image color analysis, Observers
9	RNKD RAJAPAKSE	MECHANICAL BEHAVIOR OF NANO STRUCTURES USING ATOMIC-SCALE FINITE ELEMENT METHOD (AFEM)	Latin American Journal of Solids and Structures	AFEM; INTERATOMIC POTENTIALS; LENNARD JONES POTENTIAL; MOLECULAR DYNAMICS
10	Jie Liang	Simultaneously Color-Depth Super-Resolution with Conditional Generative Adversarial Network	arXiv.org-Computer Vision and Pattern Recognition (cs.(	GAN, super-resolution, depth image, color image, image smoothing, edge detection
11	Timothy Horita, Parvan	AUTOMATIC CLOUD DETECTION AND VERIFICATION IN SATELLITE IMAGES	Int'l Conf. IP, Comp. Vision, and Pattern Recognition	clouds, shadows, identification, verification
12	Chinthaka Dinesh	Non-intrusive load monitoring under residential solar power influx	Applied Energy	Non-intrusive load monitoring under residential solar power influx

ACRL 2017: Gao and Wallace

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

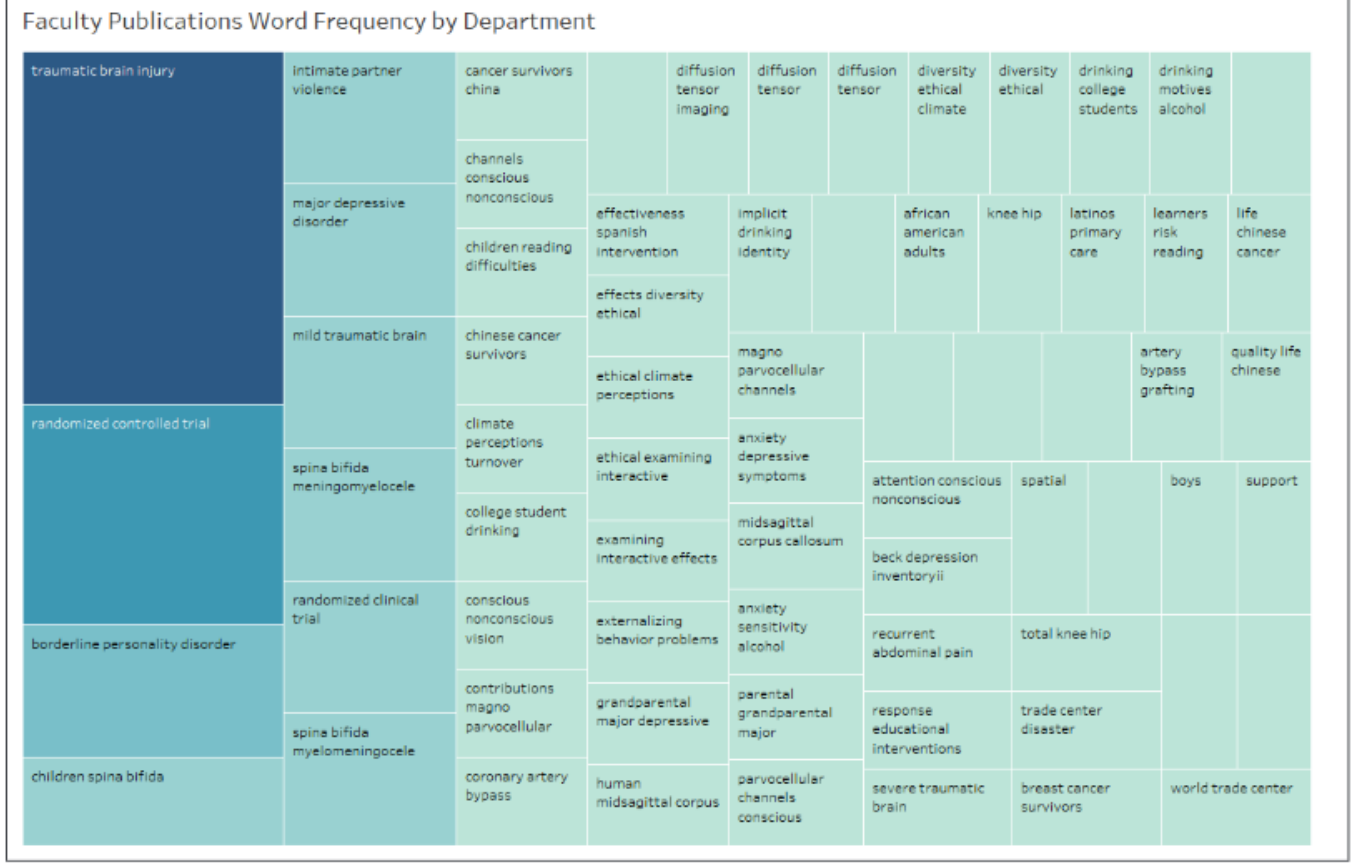
*Wenli Gao and Loretta Wallace**

Gao and Wallace's methodology

- Using Scopus, harvested citations from entire university
- Ran corpus through Mallet, a topic modeling software
- Ran topics through Tableau for visualizations
- Visualized research trends

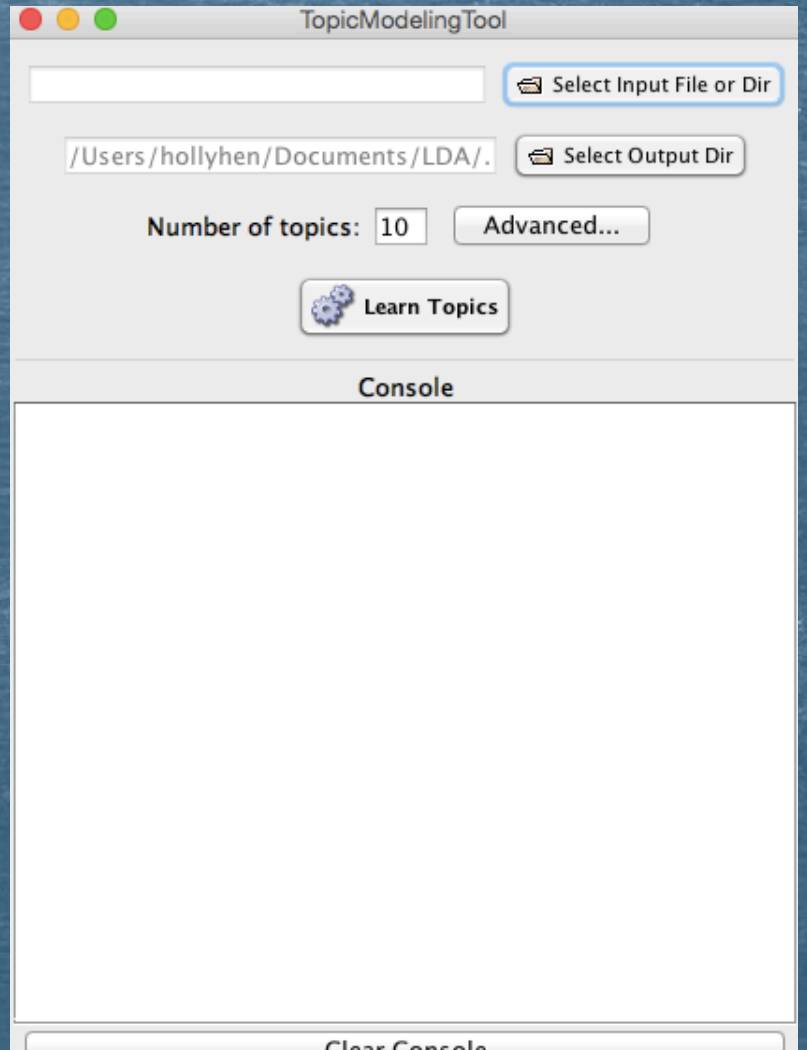
Gao & Wallace's topics (visualized)

FIGURE 1
Faculty Publications Word Frequency by Department



Holly's 1st attempt: Google Scholar/Mallet

- Downloaded Mallet (with domestic technical support, ie, husband)
- Input titles of articles into Mallet for topic analysis
- A "topic" consists of a cluster of words that frequently occur together.



Mechatronics topic model output

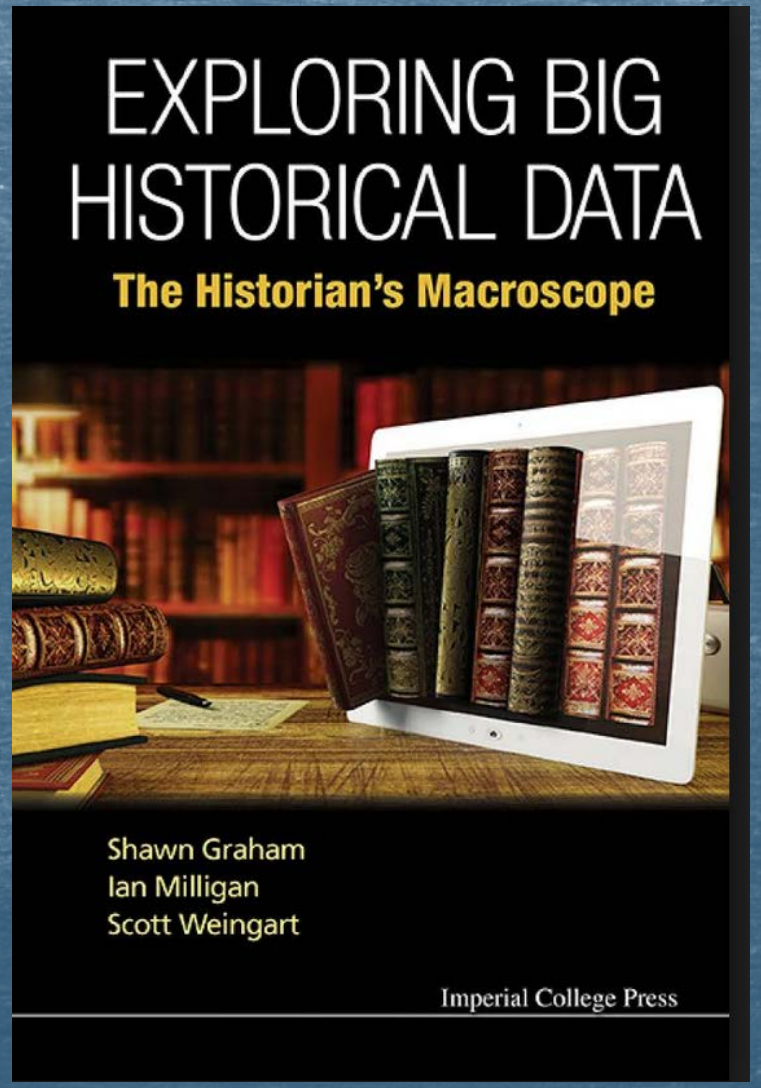
List of Topics

1. cell fuel membrane cells mechanical
2. system systems time pressure r
3. heat model performance c different
4. thermal temperature method flow through
5. on was were than surface
6. high analysis have these voltage
7. on optimization be cost based
8. design power energy control proposed
9. using algorithm can method proposed
10. hydrogen from sensor sensors which

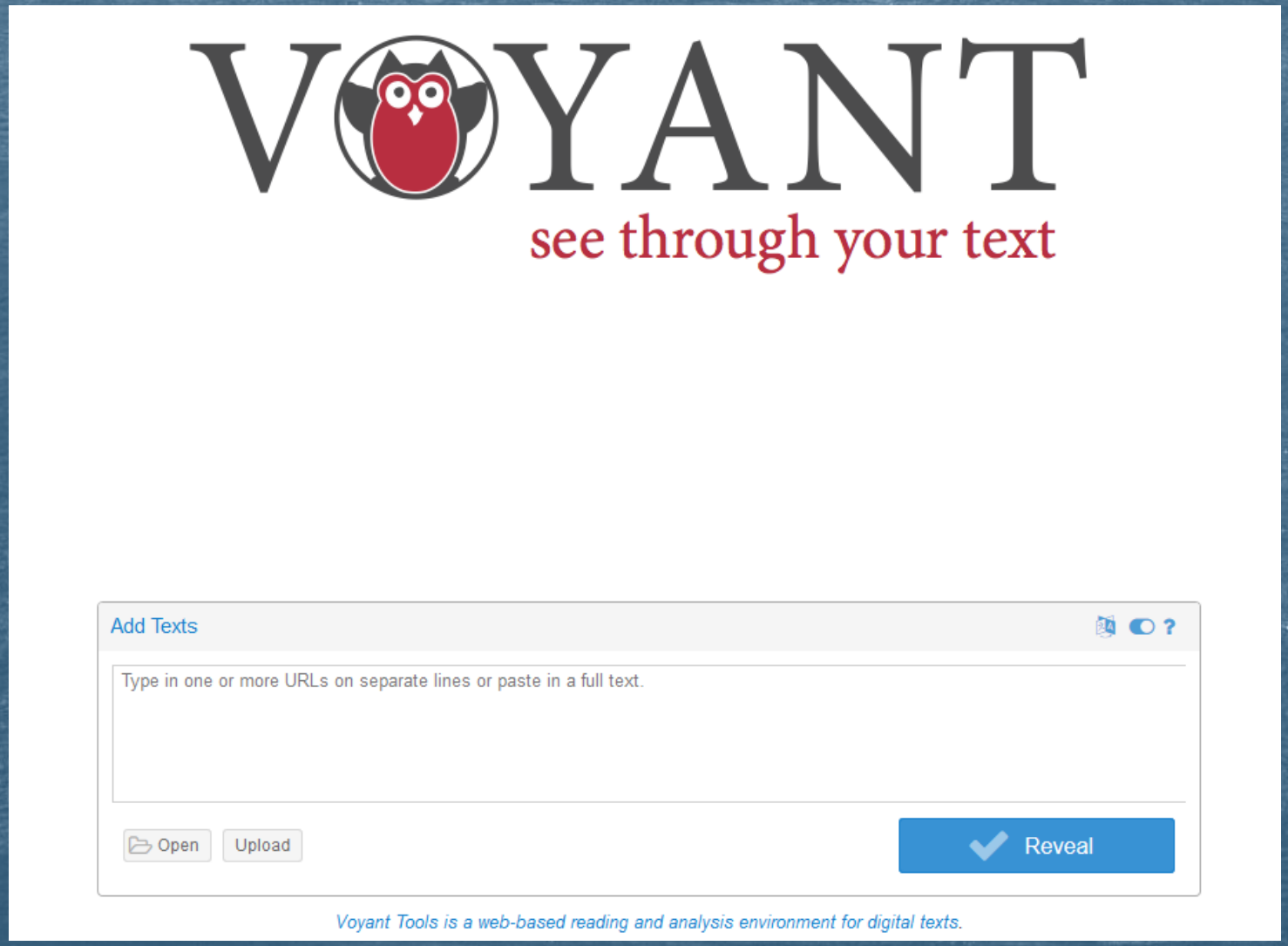
Why I failed with Mallet

- My lack of subject expertise did not allow me to understand or articulate the larger “grouping” of terms
- Subject experts are able to take these topics and label them
- Stopwords an issue

Discovery of Voyant Tools



“the gateway drug of text mining applications”



Corpus source: Web of Science

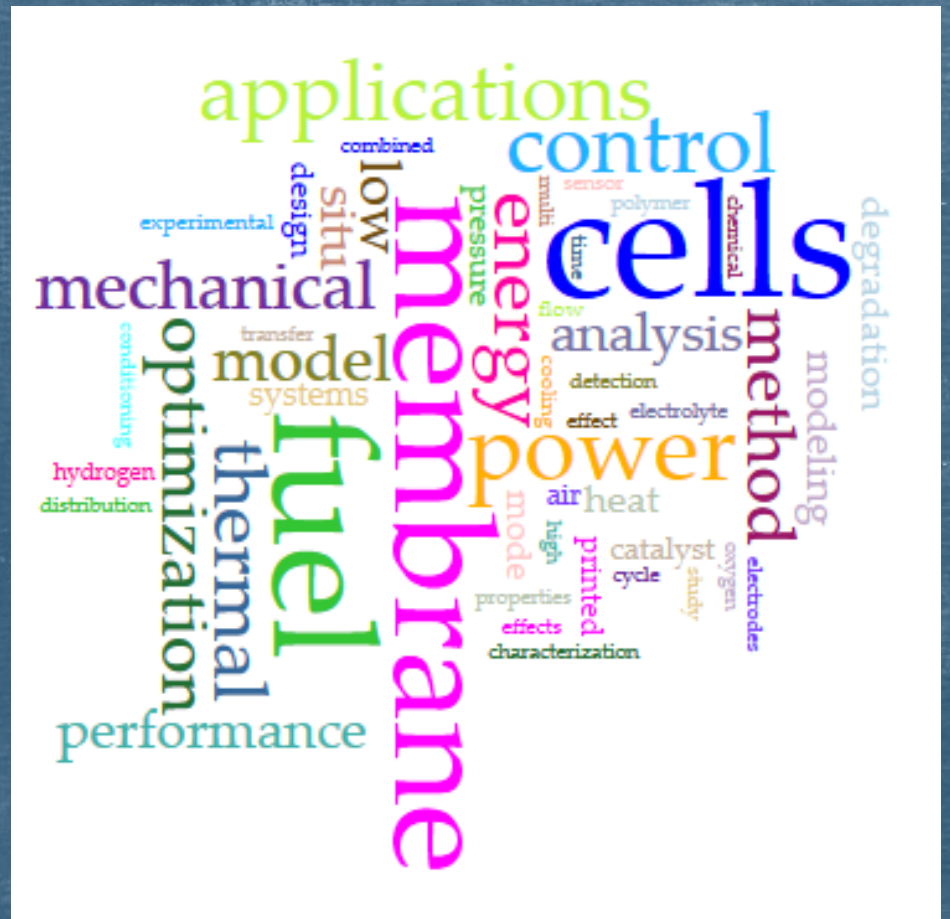
Set	Run Search **Web of Science Core Collection Search History - " eng mech"**
#3	**ADDRESS:** (Simon Fraser Univ, Mechatron Syst Engn) *DocType=All document types; Language=All languages;*
#2	**ADDRESS:** (Simon Fraser Univ, Sch Mechatron Syst Engn) *DocType=All document types; Language=All languages;*
#1	**ADDRESS:** (Simon Fraser Univ, Sch Engn Sci) *OR* **ADDRESS:** (Simon Fraser Univ, Sch Eng Sci) *DocType=All document types; Language=All languages;*

Mechatronics word bag: 234 article titles

3d failure analysis of pure mechanical and pure chemical degradation in fuel cells membrane
3d printed flow field and fixture for visualization of water distribution in fuel cells by x-ray computed tomography
3d printed inductor designs decorated with silver nano ink
3d printed stretching-dominated micro-trusses
4d in situ visualization of electrode morphology changes during accelerated degradation in fuel cells by x-ray computed to
a 3-d-printed integrated pcb-based electrochemical sensor system
a bidirectional boost converter with application to a regenerative suspension system
a capacitor voltage balancing method for cascaded h-bridge multilevel inverters with application to facts
a cascaded kalman filter-based gps/mems-imu integration for sports applications
a circuit-based approach for electro-thermal modeling of lithium-ion batteries
a comparative study of extended kalman filter and an optimal nonlinear observer for state estimation
a comparative study of the energy-saving controllers for automotive air-conditioning/refrigeration systems
a double-loop primary-side control structure for hb-led power regulation
a general active capacitor voltage regulating method for l-level m-cells n-phase flying capacitor multilevel inverter with arb voltage distribution
a general control method for multilevel converters based on knapsack
a high-performance piezoelectric vibration sensor
a low-power readout circuit design for capacitive microsensors
a magnetometer-free indoor human localization based on loosely coupled imu/uwb fusion
a metal-free and biotically degradable battery for portable single-use applications
a methodology for optimal design of a vehicle suspension system with energy regeneration capability
a methodology for reducing the filtering capacitor in low-flicker led drivers
a modified steady state method for measurement of in-plane thermal conductivity
a multi-surface sliding-mode extremum seeking controller for alternator maximum power point tracking
a new design method for the de inductance in current source converters
a novel biomechanical model-aided imu/uwb fusion for magnetometer-free lower body motion capture
a novel method for inlet duct geometry improvement of heat recovery steam generators
a novel, state-of-the-art tubular architecture for polymer electrolyte membrane fuel cells: performance enhancement, size
a paired stretchable printed sensor system for ambulatory blood pressure monitoring
a portable system for estimation of chemical oxygen demand in wastewater using ultraviolet-visible spectroscopy
a preliminary investigation into the design of pressure cushions and their potential applications for forearm robotic orthose
a preliminary investigation on the utility of temporal features of force myography in the two-class problem of grasp vs. no- the presence of upper-extremity movements

Live demo

Mechatronics cirrus



Engineering cirrus

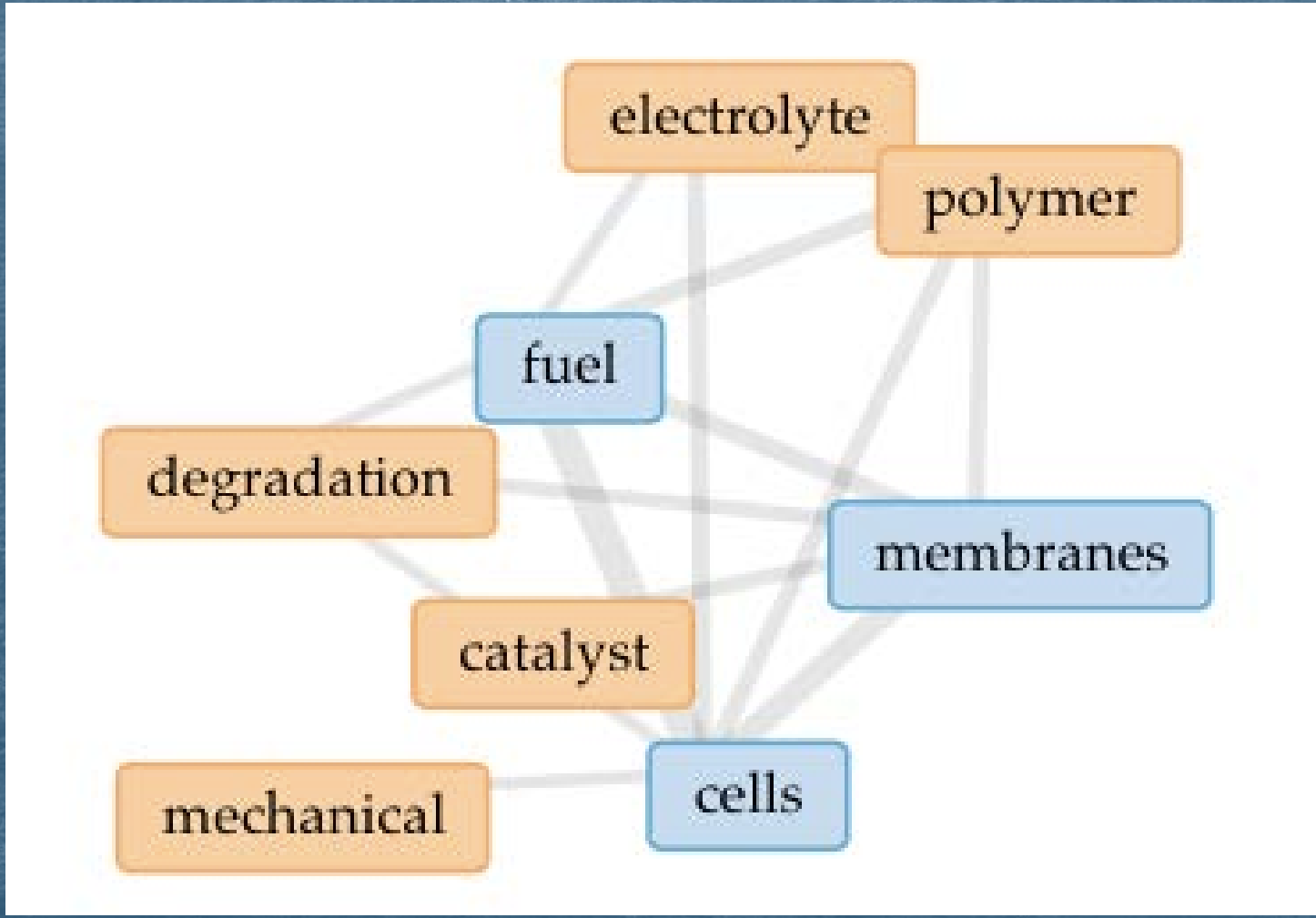
Corpus summary statement: Mechatronics

This corpus has 1 document with 3,257 total words and 1,134 unique word forms.

Vocabulary Density: 0.348

Most frequent words in the corpus: cells (46); fuel (40); membrane (32); control (20); power (20); applications (19); energy (17); thermal (17); optimization (16); mechanical (15); method (15); model (14); analysis (13); low (13); performance (13); situ (13); degradation (11); heat (11); modeling (11); air (10); catalyst (10); design (10); mode (10); pressure (10); printed (10); systems (10); conditioning (9); cooling (9); experimental (9); flow (9); hydrogen (9); time (9); characterization (8); chemical (8); cycle (8); detection (8); effect (8); electrolyte (8); high (8); multi (8); oxygen (8); polymer (8); properties (8); sensor (8); study (8); transfer (8); combined (7); distribution (7); effects (7); electrodes (7); exchange (7); fatigue (7); investigation (7); layers (7); micro (7); point (7); porous (7); real (7); refrigeration (7)

Colocates graph



Keyword in context

		Left	Term	Right ↑
⊞	1)	for measurement of in-plane	thermal	conductivity a multi-surface sliding
⊞	1)	eulerian and lagrangian approaches effective	thermal	conductivity modeling of consolidated sorp…
⊞	1)	a/c system for trucks	thermal	conductivity of catalyst layer of
⊞	1)	cells: part 1-experimental study	thermal	conductivity of catalyst layer of
⊞	1)	cells: part 2-analytical modeling	thermal	conductivity of microporous layers: analytical
⊞	1)	plane source method for measuring	thermal	conductivity of thin films: deconvoluting
⊞	1)	conductivity of thin films: deconvoluting	thermal	contact resistance an integrated microfluidic
⊞	1)	refrigeration systems analytical investigatio…	thermal	contact resistance (tcr) behavior under
⊞	1)	electronics counter-intuitive reduction of	thermal	contact resistance with porosity: a
⊞	1)	inverse method for calculation of	thermal	inertia and heat gain in
⊞	1)	tcr) behavior under time-dependent	thermal	load application of carbon nanotube
⊞	1)	using iec 61850 real-time	thermal	load calculation by automatic estimation
⊞	1)	for real-time calculation of	thermal	loads in hvac-r applications
⊞	1)	circuit-based approach for electro-	thermal	modeling of lithium-ion batteries

Remember to shift the lens

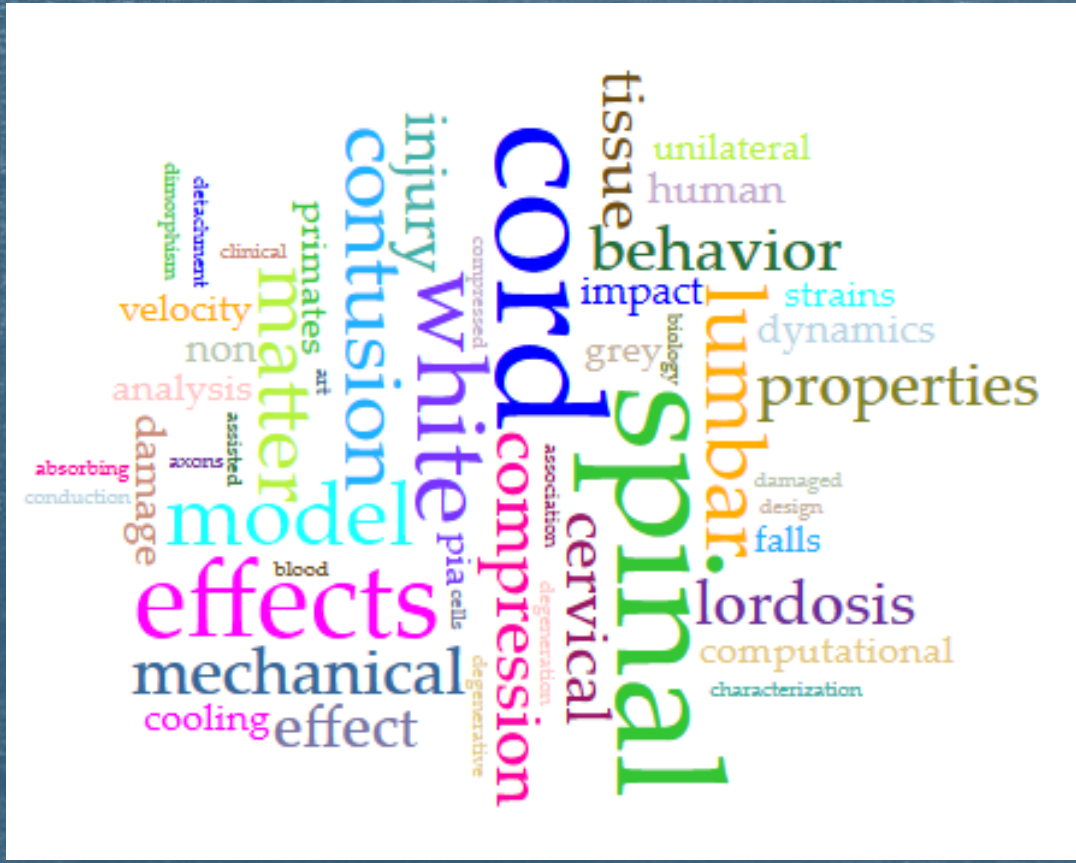
- Dr Carolyn Sparrey
- Biomechatronics research group

Research interests (website)

- Mechanical characterization of biological tissues
- Phenomenological constitutive models of tissues.
- Patient specific technologies
- Modeling and simulation of human injury and disease
- Accident reconstruction and injury analysis
- Design and evaluation of injury prevention devices

Research interests/Research output

Cirrus view



Terms view

	Term	Count
1	cord	10
2	spinal	10
3	effects	6
4	white	5
5	contusion	4
6	lumbar	4
7	matter	4
8	model	4
9	behavior	3
10	cervical	3
11	compression	3
12	effect	3
13	injury	3
14	lordosis	3
15	mechanical	3
16	properties	3
17	tissue	3

Big Data approach to research output

Graham: “big data is simply more data that you could conceivably read yourself in a reasonable amount of time – or, even more inclusively – information that requires computational intervention to make new sense of it.”

- Voyant Tools rearranges and decontextualizes word frequencies, which helps with
  - 1. learning vocabulary
  - 2. macroscope view

Macroscope metaphor

- As opposed to close reading (microscope)
- A way to pull back, find structures within the data

Franco Moretti: “Distant Reading” theory

- Reading more / reading closely not the solution to an understanding of world literature
  - Canon is small (less than 1% of all texts)
  - Nobody can read everything!

- Distant reading “allows you to focus on units that are much smaller or larger than the text: devices, themes, tropes, or genres or systems.”

Macroscopes and metadata

- LIS workers not close readers, anyway
- Experts in Metadata
- Difference: we usually work at the object (document) level rather than the corpus
- Corpus analysis flips our practice: “the sources speak to you” (Graham, Milligan and Weingart”

Mistrust of text mining

- Ronald Hagler: Librarians “shunned” natural language access points since the 19th century in favour of controlled vocabulary
- Mechatronics wordbag contains both “3-D” AND “three-dimensional.”
- Must acknowledge these limitations

Rebuttal: Scientific language

- Ronald Hagler: “the precision of much of [scientists’] terminology and writing gives their natural language many of the characteristics of a controlled vocabulary”
- Engineers not prone to puns or wordplay

Technical Papers

Downloaded 204 times

Numerical Collapse Analysis of Tsuyagawa Bridge Damaged by Tohoku Tsunami

Hamed Salem; Suzan Mohssen; Yuto Nishikiori; and Akira Hosoda

FULL TEXT DOWNLOAD TOOLS SHARE

NOT

[ALBUM COVER, SIMON AND GARFUNKLE'S BRIDGE OVER TROUBLED WATER]

Further Use for Voyant Tools

- “Reveal” other corpuses, including journal titles, curriculum documents, theses
- Map word frequencies to controlled vocabulary (eg, collections profiles or broader categories)
- Upload a “word bag” when making selections VERY HELPFUL for technical subjects

Conclusion

- Voyant Tools provides "macroscope" on research output
- Easy entry into word and phrase frequencies and networks
- Assists with an overview of the department not obtainable through the website OR the articles themselves

Bibliography

Gao, W., & Wallace, L. (2017). Data Mining, Visualizing, and Analyzing Faculty Thematic Relationships for Research Support and Collection Analysis. In *ACRL 2017 Conference Proceedings* (pp. 171–178). Baltimore: ALA.

Graham, S., Milligan, I., & Weingart, S. (2015). *Exploring Big Historical Data: The Historian's Macroscope*. London: World Scientific Publishing.

Hagler, R, & Simmons, P. (1991). *The bibliographic record and information technology.* Chicago: ALA.

Maness, J. M. (2016). Engineering and Applied Science Librarianship. In K. Sobel (Ed.), *Mastering Subject Specialties: Practical Advice from the Field: Practical Advice from the Field* (pp. 31–38). ABC-CLIO.

Moretti, F. (2013). *Distant reading*. London : Verso.

Tobin Cataldo, T., Tennant, M. R., Sherwill-Navarro, P., & Jesano, R. (2006). Subject specialization in a liaison librarian program. *Journal of the Medical Library Association*, *94*(4), 446–448.

Questions & Discussion