

Analyzing Trends in AAAR Topics over the Past Two Decades

Brandon Beattie¹

Supraja Gurajala¹ and Suresh Dhaniyala²

¹Computer Science, SUNY Potsdam

²Mechanical and Aerospace, Clarkson University

Acknowledgment:

Data source: Donald Dabdub and Susanne Hering

Gain insight into

Goal

- Prevalent topics within conference papers
- Trends of topics over the years
- Change in vocabulary

Methods and Tools

Data Collection:

MySQL, Excel, PyPDF2

Document Preprocessing:

NLTK, Gensim

Corpus & Dictionary:

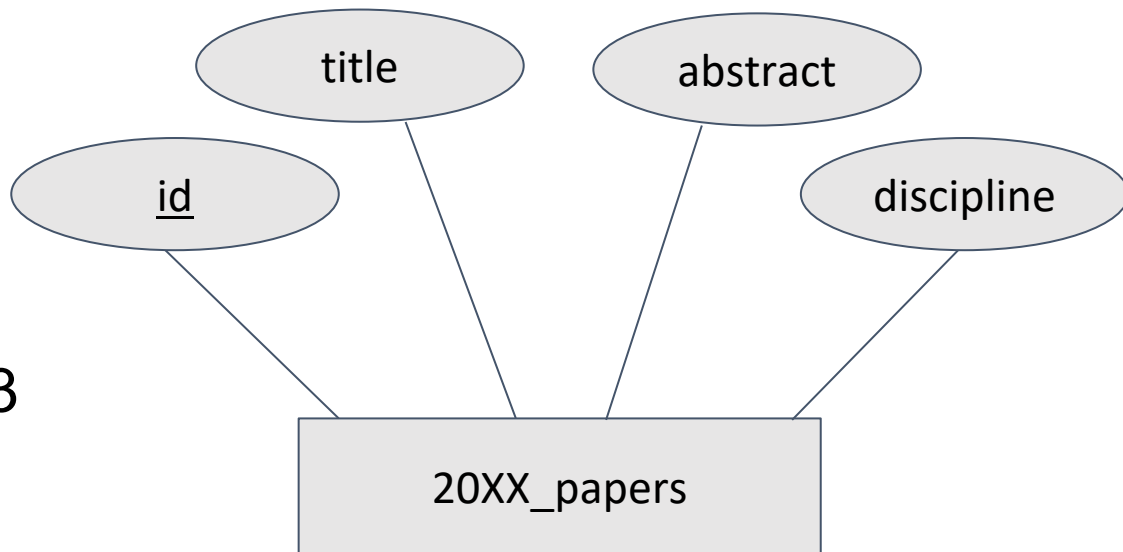
Gensim

Model to extract topics:

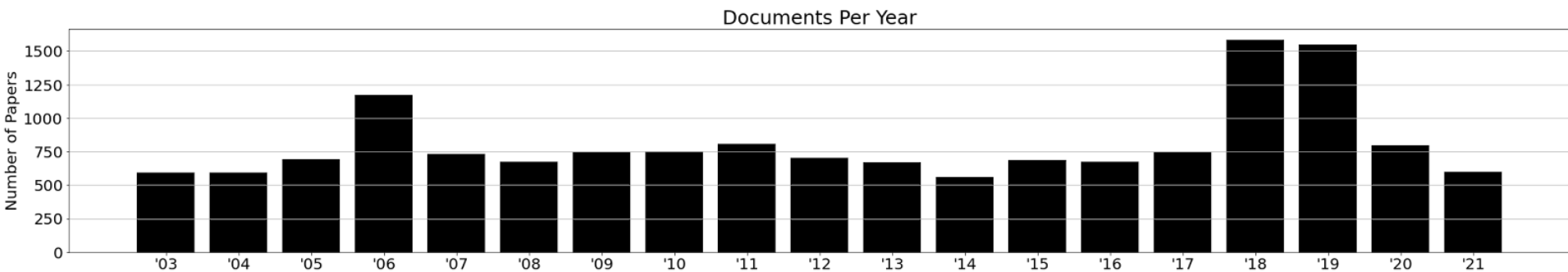
Gensim

Data Collection

- Given SQL, Excel, & PDF files from AAAR conferences from 2003 to 2021
- Consolidated in MySQL

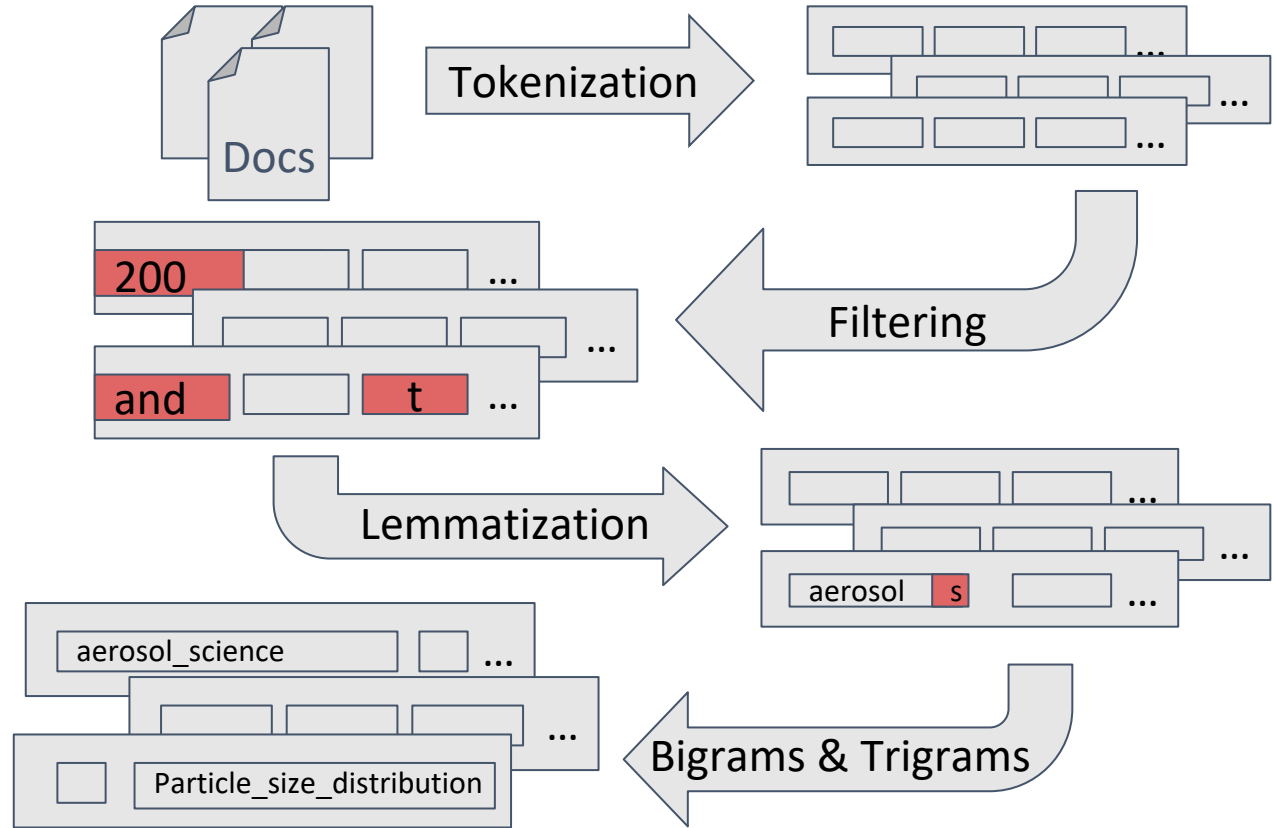


Abstracts over the years



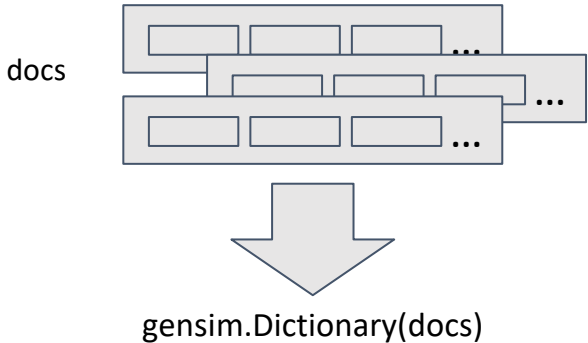
Data Pre-Processing

- Tokenization
- Initial Filtering
 - Stopwords
 - Length < 2
 - Numbers
- Lemmatization
- Add Ngrams

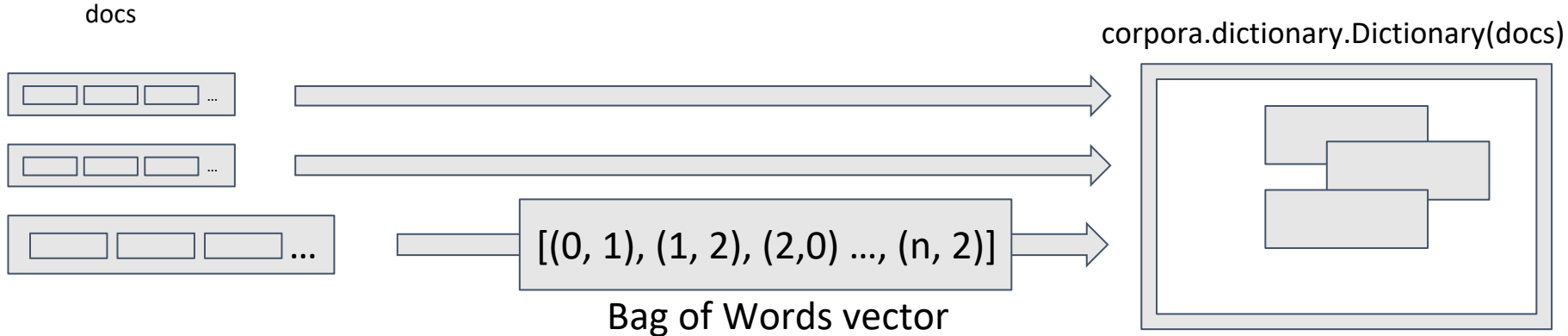


Corpus & Dictionary

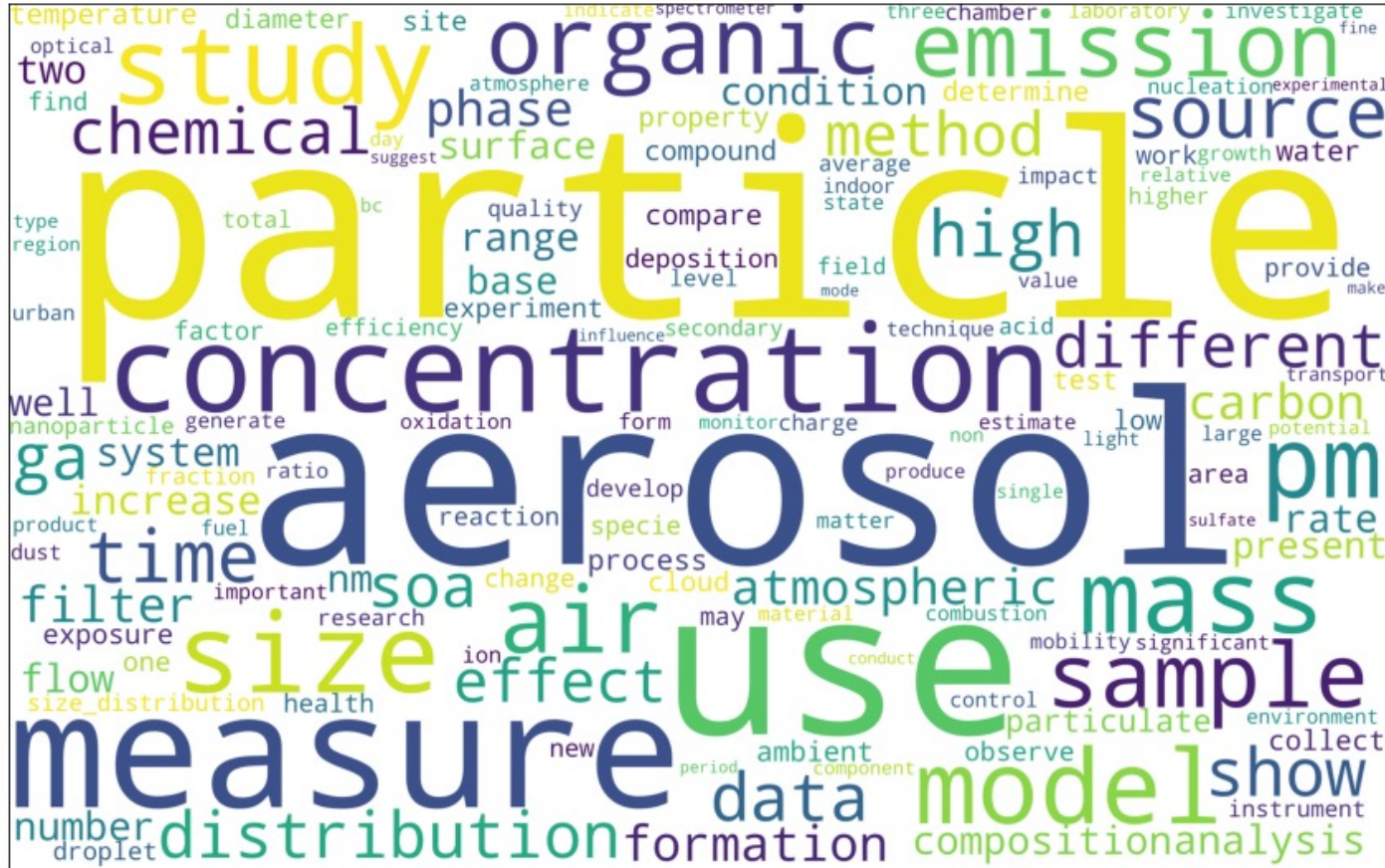
- Dictionary Generation
 - Filtering by document involvement
- Corpus Conversion



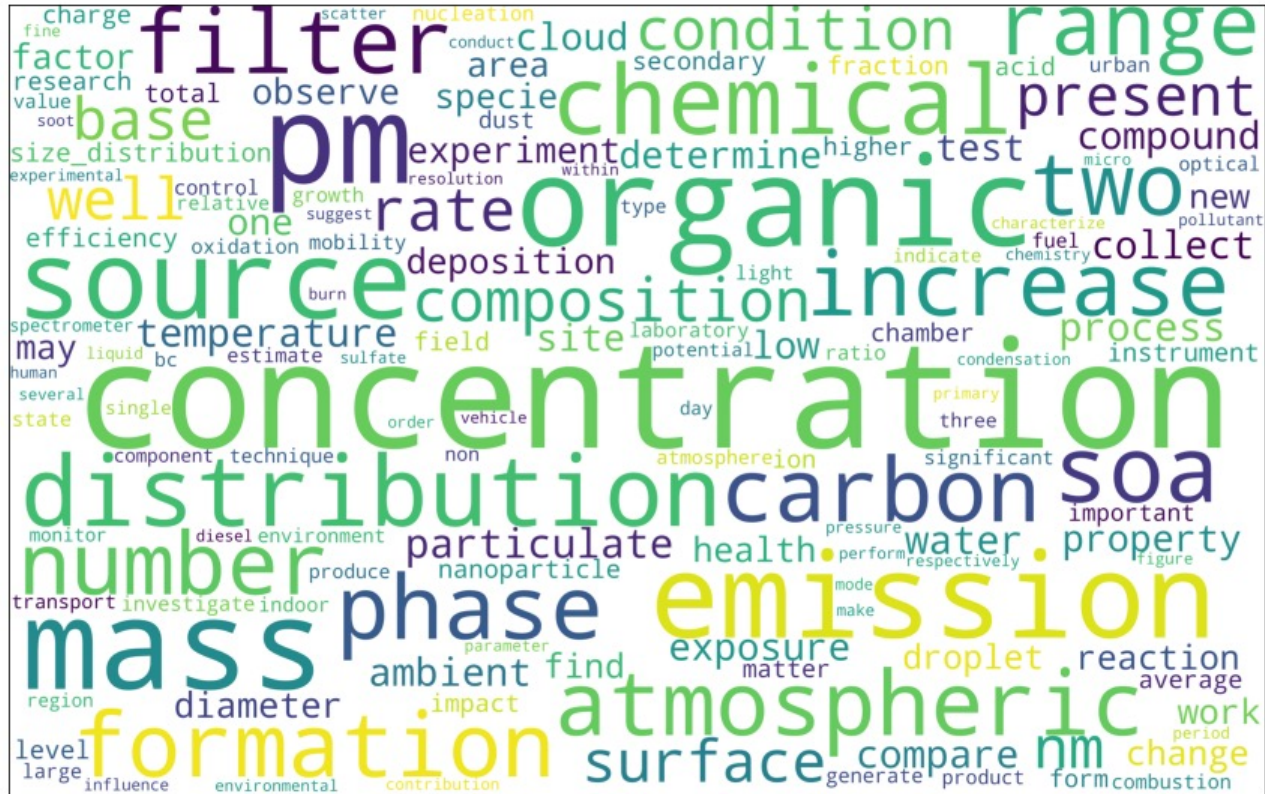
0 "aerosol"	1 "science"	2 "constantly"	3 "evolving"
4 "model"	5 "___"	6 "___"	...



Overall Word Cloud



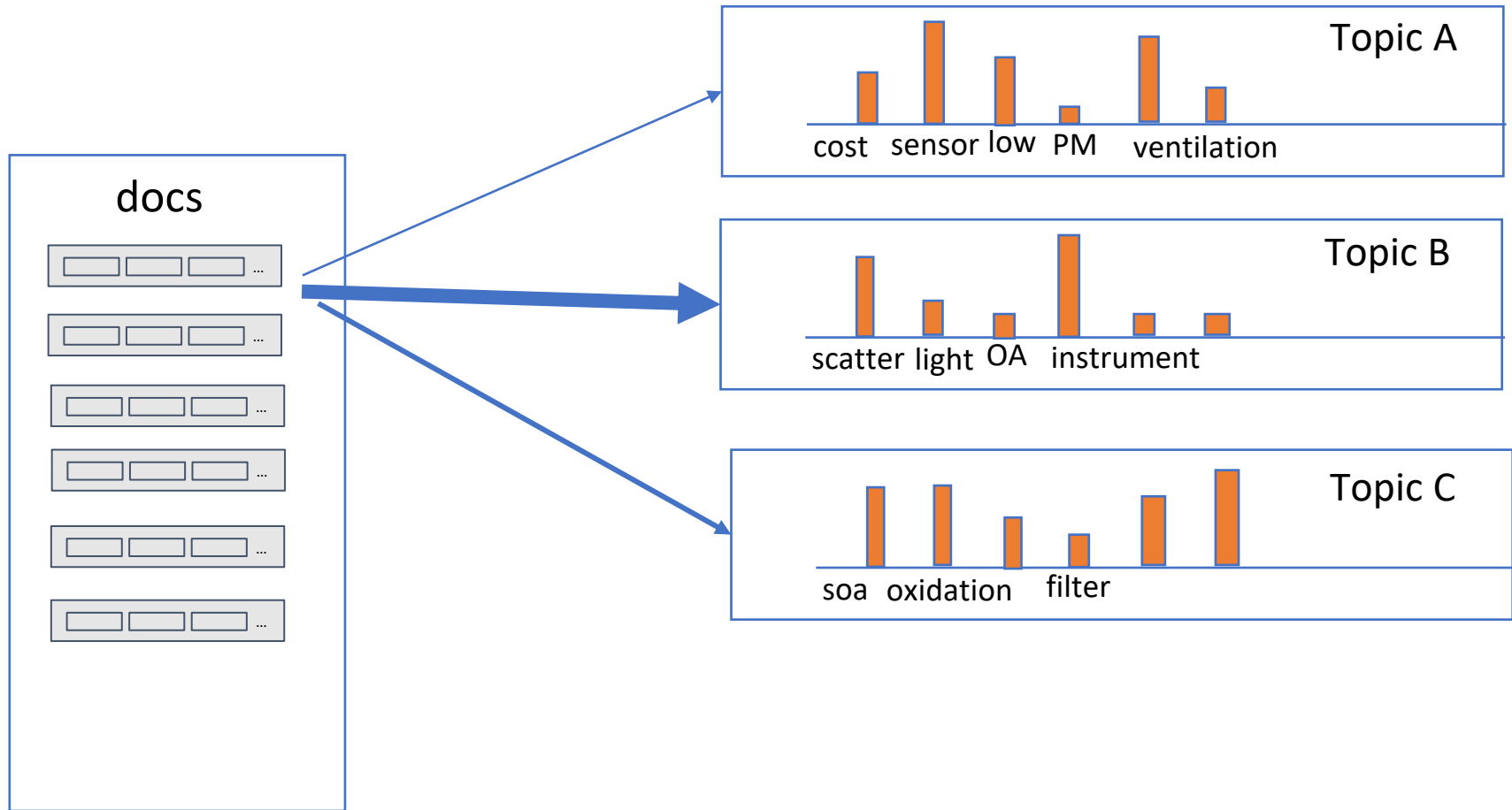
Overall Word Cloud - After Filtering

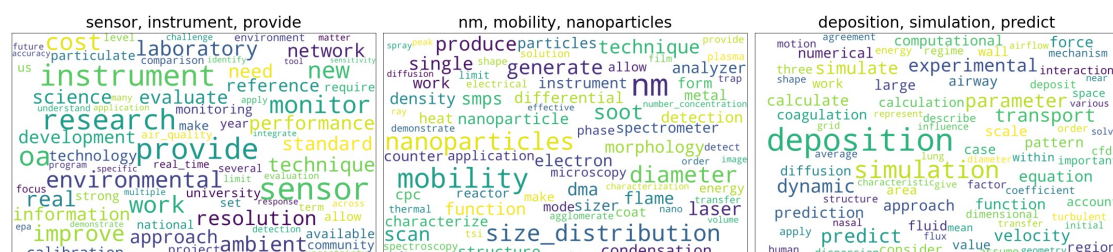


Topic Model

- Latent Dirichlet Allocation (LDA)
 - Unsupervised learning algorithm
- Choosing Parameters
 - Number of Topics - 12
 - Iterations & Passes

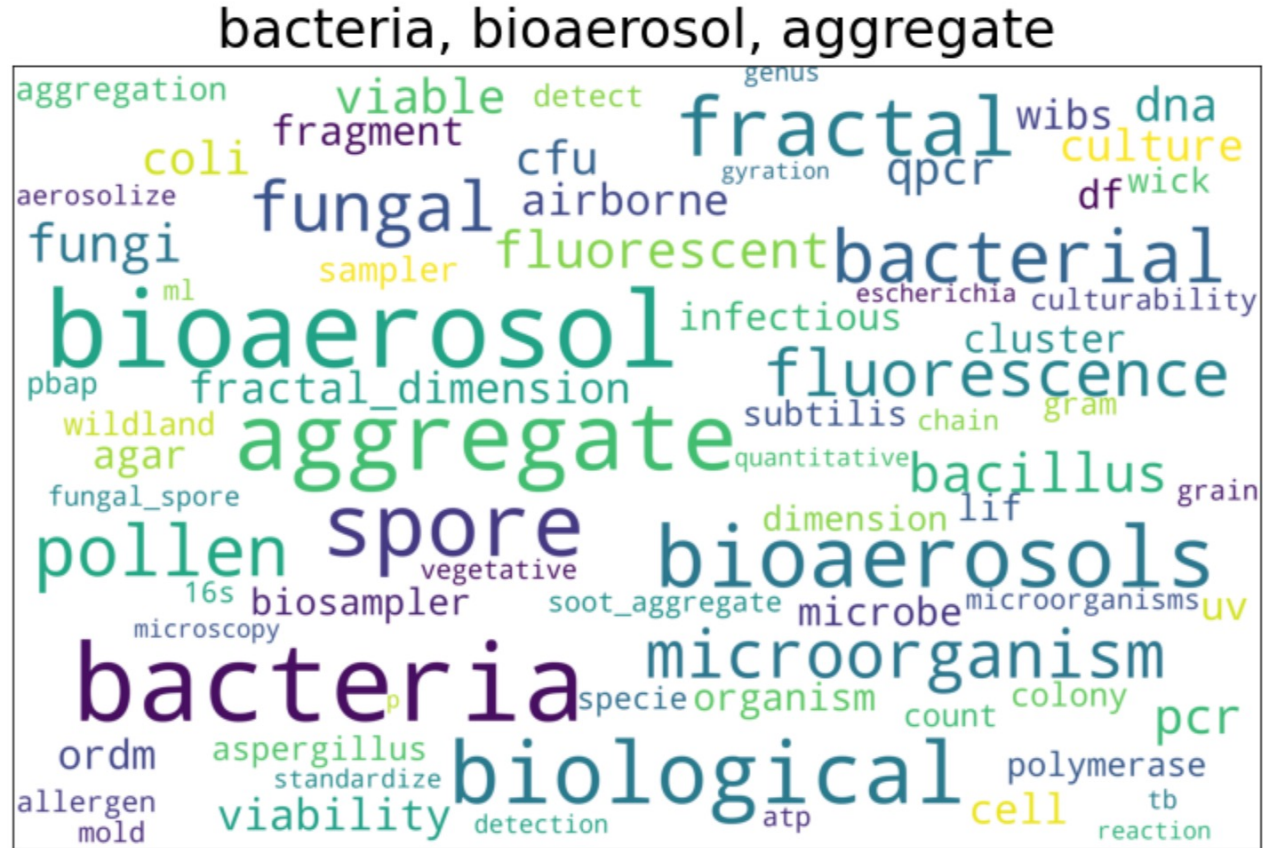
LDA

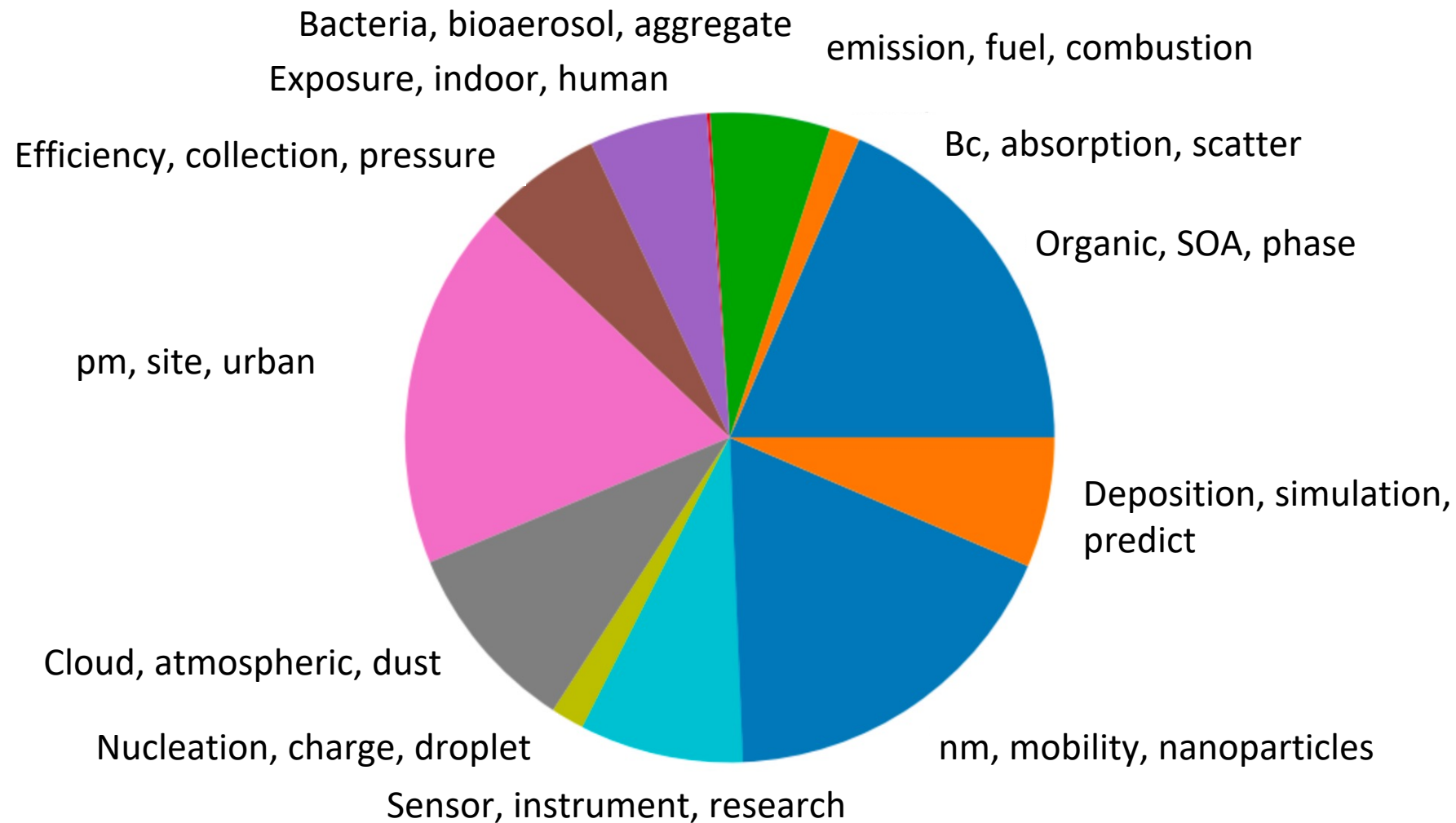


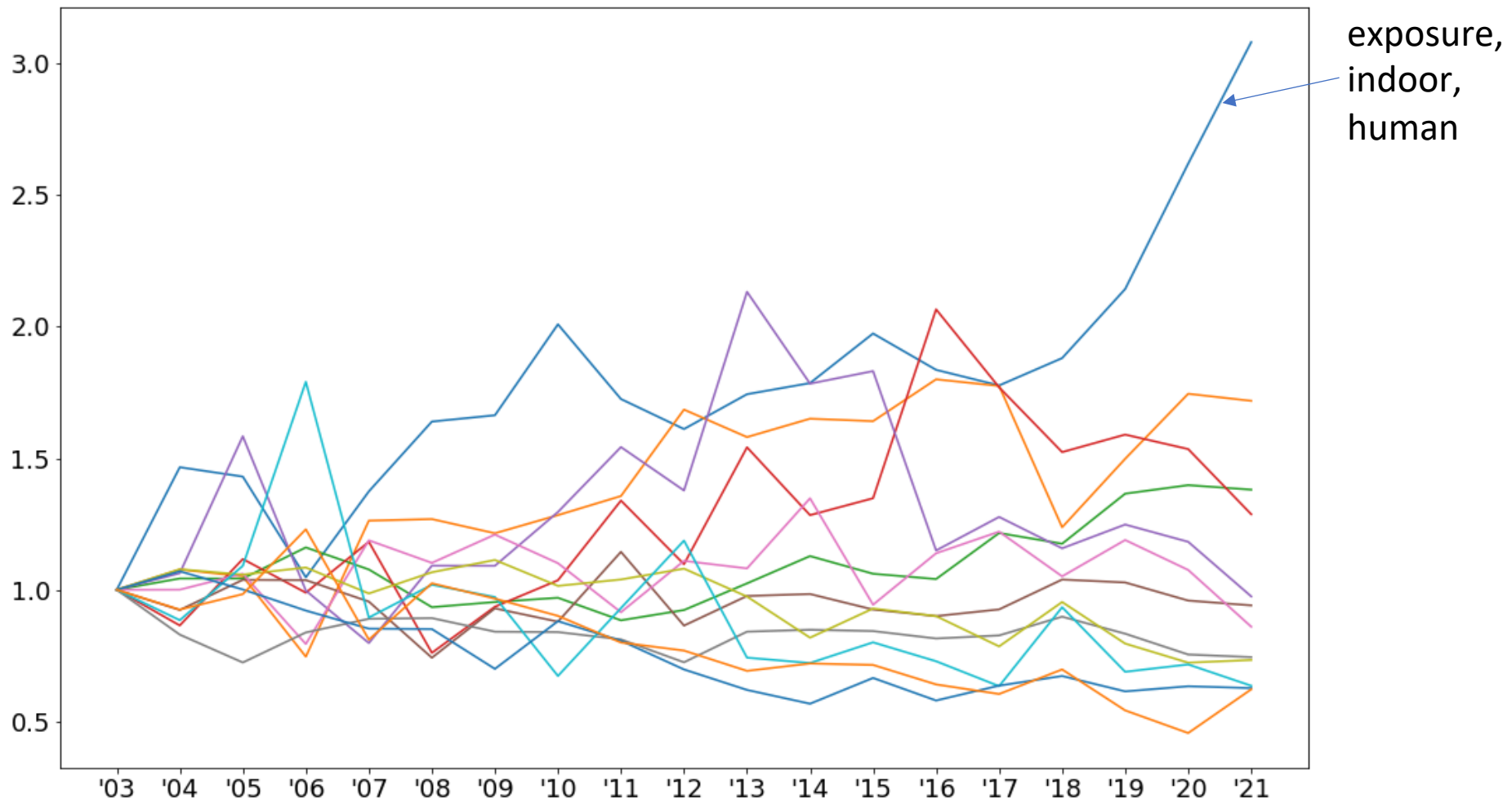


Sample topics extracted

- Words within topics sized by their frequency of occurrence
- Automated topic names based on top three words

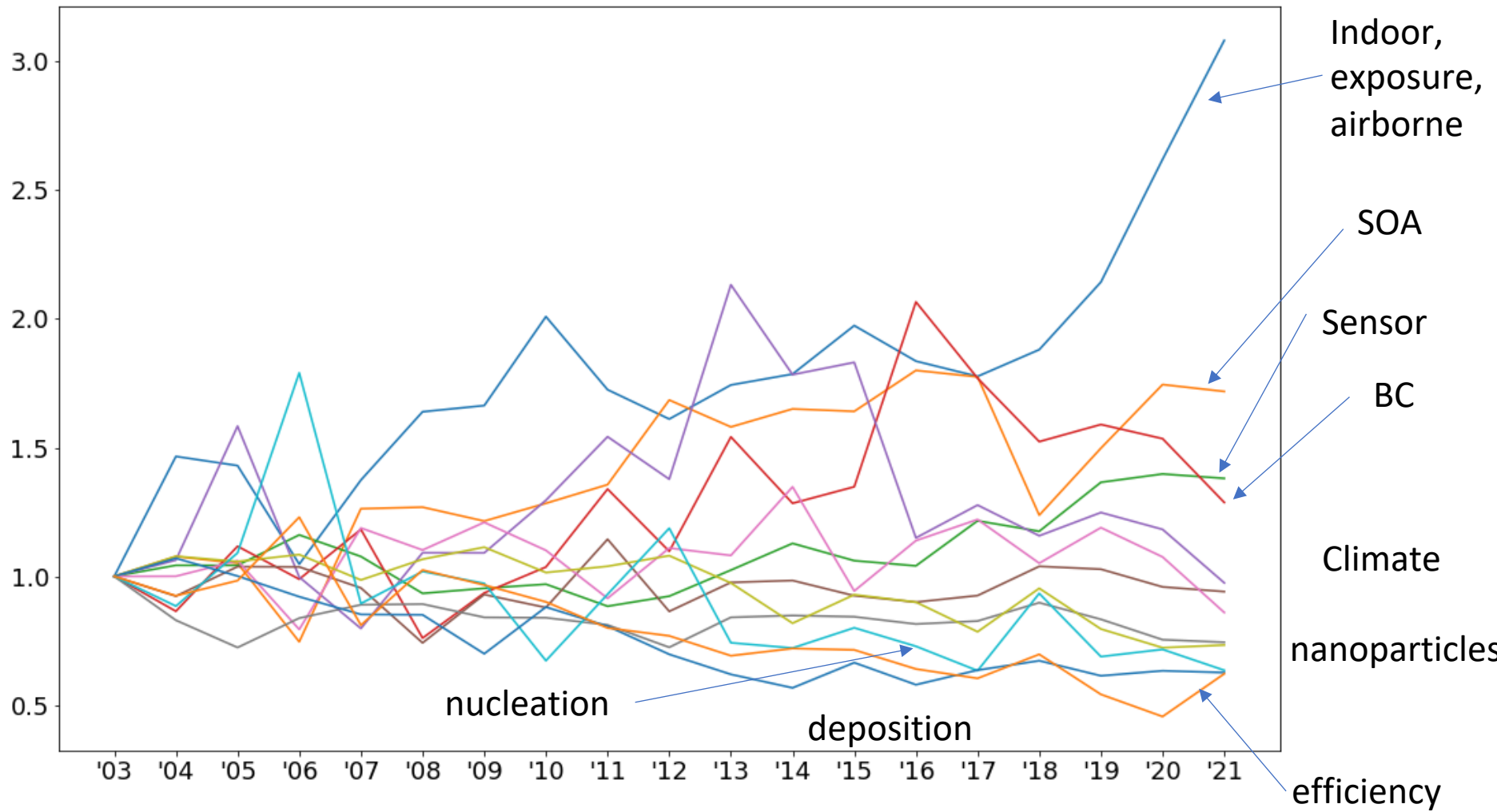






exposure, indoor, human

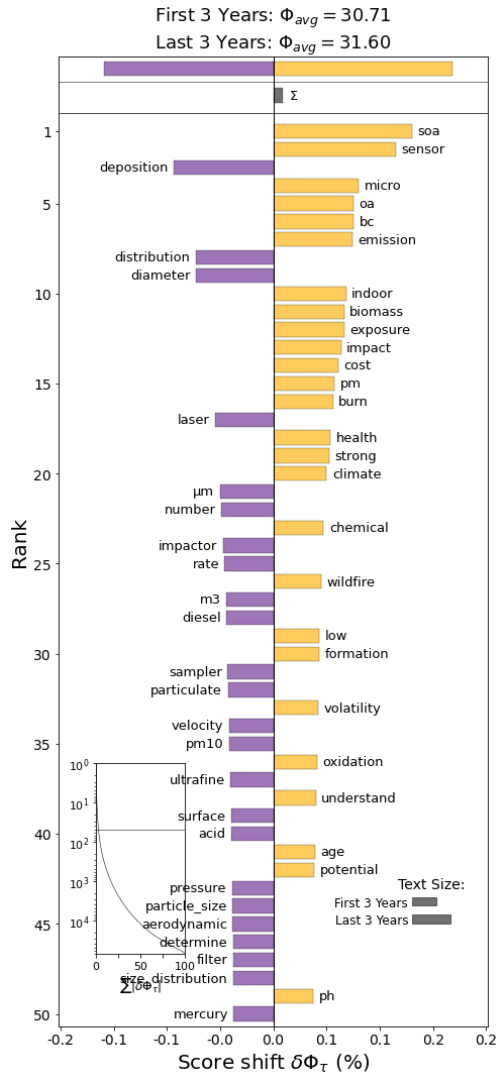




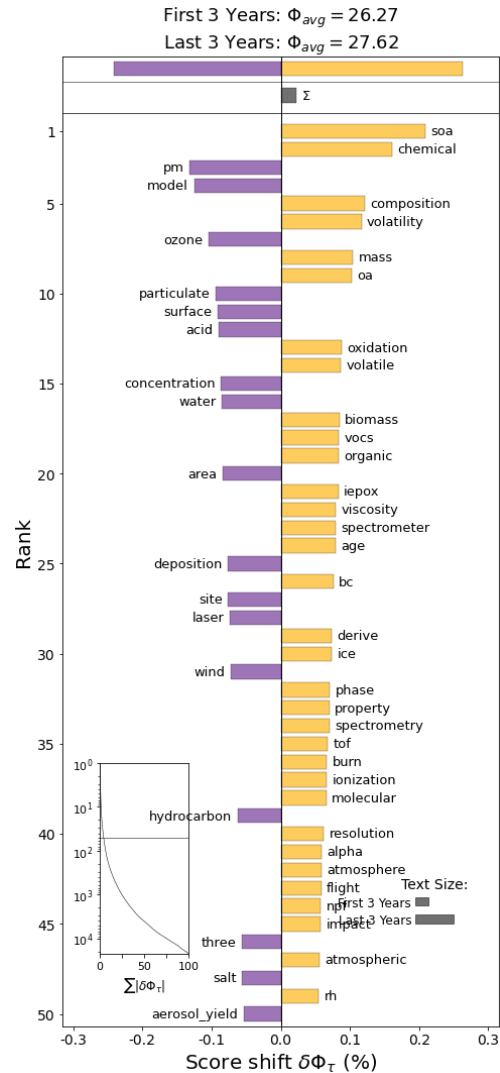
[illegible]

Change in vocabulary

Entire Collection



SOA Topic



Summary

- Automated topics extracted demonstrate the breadth of AAAR research
- AAAR topics have evolved over the last 18 years
 - As expected the pandemic has resulted in a spurt of virus-related work
 - Steady, consistent growth of SOA related presentations
 - Some topics are in decline – filtration, lung deposition, nucleation, etc
- Automated analysis can possibly aid in improving AAAR conference organization