Who, what, when, where, and why?

• Why the change to automated indexing?
• When was indexing automated?
• What algorithm is used for automated indexing?
• Who reviews automated indexing?
• Where is algorithm development going, and where can I find more information or report problems?
Why the change to automated indexing? (problem)

Problem: human indexing did not scale

- 1,000,000+ articles published
- 600,000 article backlog
- Many months to index
Solution: MEDLINE 2022 Initiative

Well-performing algorithm +

MTI NLM Medical Text Indexer
Providing Indexing Assistance
Since 2002

Biomedical Literature → MTI → MTIFL → MeSH Suggestions → MTIA

Human quality assurance
The impact of changing to automated indexing.
When was indexing automated?

April 2022!

2002

MTI indexing suggestions

2012

MTI first line (MTIFL)

2022

MTI comments (MTIC)

MTI Review

NIH National Library of Medicine
What algorithm is used for automated indexing?
MTIA Workflow

1. Process title and abstract
2. Map to MeSH; ML component
3. PubMed similar articles
4. Rank, filter, boost
5. Automated MeSH indexing
MTIA Performance

MTIA performance

- Overall
- MeSH terms
- Check Tags
- All SCRs
- Protein SCRs
- Publication Types
- Subheadings

Graph showing precision and recall for different categories.
Who reviews automated indexing?

Team of curators with domain expertise

Daily review of automation results

- Metaphor
- Genes/proteins
- Specific PTs

Random set of 10% of daily load
PMID 35819125 (Expert Rev Respir Med)

TI - Risk factors associated with mortality in hypersensitivity pneumonitis: a meta-analysis.

AB - BACKGROUND: Hypersensitivity pneumonitis (HP) related deaths have increased substantially in recent years. It is important to identify the risk factors of HP significantly associated with mortality to ensure close patient monitoring and assess disease progression. RESEARCH DESIGN AND METHODS: Extensive literature search was conducted in accordance with the PRISMA checklist. Literature search of PubMed, Embase, and Cochrane Library database between January 2009 and April 2021 using the terms 'hypersensitivity pneumonitis', 'hazard ratio', and 'mortality' identified 325 articles. A total of 22 independent original studies focusing on mortality of HP patients were assessed. RESULTS: This systematic review and meta-analysis suggests that increased age, male sex, honeycombing, and traction bronchiectasis patterns on high-resolution computed tomography (HRCT) images are the major mortality-related risk factors of patients with HP. In case of chronic HP, antigen exposure appeared to be an additional risk factor. CONCLUSIONS: The clinico-radiological risk factors of mortality identified for HP will enable effective and close monitoring of patients, prognostication, and guide toward appropriate management decisions. However, association between the type of antigen and mortality remains to be explored.
Where is algorithm development going?

- Expanded machine learning
- Continuing refinement
- Retraining sets
- Chemical algorithm development
Where can I find more information or report problems?

**MEDLINE 2022: A Five-Year Development Plan**

**MEDLINE 2022: Transition to Automated Indexing**

Incorporating Values for Indexing Method in MEDLINE/PubMed XML

Frequently Asked Questions about Indexing for MEDLINE
What are the main takeaways?

- Indexing backlog eliminated
- MEDLINE indexing within 1-2 days
- Well-performing algorithm
- Human quality assurance