Comparing the National Library of Medicine (NLM)’s Medical Text Indexer (MTI) to Human Indexing: A Pilot Study

Eileen Chen, MLIS
School of Information, University of British Columbia

Dr, Julia Bullard, MA, MLIS, PhD
School of Information, University of British Columbia

Dean Giustini, MLS, MEd
UBC Biomedical Branch Library, Vancouver Hospital
Background on the Medical Text Indexer (MTI):

MTI was developed in 2002 by the National Library of Medicine (NLM) & Lister Hill National Centre for Biomedical Communications.

There have been three versions:

- **MTI (2002)**: term recommender for human indexers
- **MTI First Line (2011-2021)**: semi-automated MTI assisted by human indexers
- **MTI-Auto (2022)**: fully automated indexer (extent of human curation unknown)

For more information (abstract, data, glossary, etc) on this project, see: https://osf.io/4k69q/
MTI-Auto (2022):

- Most recent version, but not publicly-available for testing
- *Pattern-based indexing* based on titles & abstracts of papers
- Machine learning used for applying subheadings
- Human reviewers perform quality assurance reviews for *‘select citations’* ...

For more information on this project, see: https://osf.io/4k69g/
Aims:

- To compare MTI with human indexing...
- Evaluate indexing quality in high vs. low-impact biomedical journals indexed in Medline (PubMed)
- Identify MTI errors, and anomalies in assigning MeSH terms & check tags

For more information on this project, see: https://osf.io/4k69q/
Method used for sample:

- **Selected** 20 biomedical articles published in year 2000 (before MTI was created)
- **Identified** key journals from the ‘Abridged Index Medicus’ (AIM) = a journal subset of Medline (PubMed)
  - Of 120+ AIM journals, 10 with the highest 2020 Journal Impact Factor (JIF) AND 10 with the lowest JIF were chosen (N=20)
- **Excluded** articles without abstracts or MeSH indexing
Interactive MTI Tool

...is a free online tool provided by the NLM. The version of the MTI available is the MTIFL (retired by the NLM in 2021)...
Results
Assigned Index Terms – Mean #?

- MTI and humans created more index terms for high-JIF group than low-JIF group.
- Difference was greater for MTI (6.4 terms).
- Journals with most MTI terms?
  - Lancet (26), JAMA (21), Blood (21), Annals of Internal Medicine (21).
- Journals with least MTI terms?

<table>
<thead>
<tr>
<th></th>
<th>MTI assigned:</th>
<th>Human assigned:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10 JIF journals:</td>
<td>16.6 terms</td>
<td>13.5</td>
</tr>
<tr>
<td>Lowest 10 JIF journals:</td>
<td>10.2 terms</td>
<td>11.2</td>
</tr>
<tr>
<td>Difference?</td>
<td>6.4 terms</td>
<td>2.3</td>
</tr>
</tbody>
</table>
Main Headings

- Of a total of 174 main headings used by humans for 20 articles, MTI included:
  - 80 in JTF list
  - 92 in Full Listing
  - Missed 2 altogether.
- In 19 instances, the MTI used an acceptable synonym to a human-indexed term.

The recall rate for relevant terms is high in the Full Listing, but many relevant terms are not ranked highly enough.
Case Study: MTI vs. Human Terms

- The MTI missed several major headings
- The word “attention” was misinterpreted literally:

“The concept of nursing practice models [...] has attracted the attention of nursing administrators in the last decade...” (Upenieks 2000)

<table>
<thead>
<tr>
<th>Shared Terms</th>
<th>MTI Only</th>
<th>Human Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans (0); Job</td>
<td>Social Responsibility (2);</td>
<td>Models, Nursing* (5); Nursing* (8); Outcome</td>
</tr>
<tr>
<td>Satisfaction* (1)</td>
<td>Climate Change (3); Attention (4)</td>
<td>Assessment, Health Care* (31); United States (53)</td>
</tr>
</tbody>
</table>

Italics = check tags; * = human-indexed major heading; () = MTI rankings in Full Listing
Check Tag (Age, Sex, Species) Coverage

- Of 72 check tags used by humans across sample (N=20) articles, MTI shared 38.
- Of remainder, 30 appeared in Full Listing, and 4 were missed altogether.
- MTI used 5 check tags not used by humans, 4 of which were appropriate choices.

The check tag “Aged” was missed in 4 instances.
Sex Check Tags

*Male (check tag)* was ranked higher in all 6 instances in which *male/female* check tags were used.

Why is there such a **bias** in the sample?

<table>
<thead>
<tr>
<th>Article</th>
<th>Male</th>
<th>Female</th>
<th>Difference (F - M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0*</td>
<td>3rd*</td>
<td>3 ranking places</td>
</tr>
<tr>
<td>2</td>
<td>5*</td>
<td>8*</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>1*</td>
<td>4*</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>12*</td>
<td>60*</td>
<td>48</td>
</tr>
<tr>
<td>20</td>
<td>9*</td>
<td>70*</td>
<td>61</td>
</tr>
<tr>
<td>Mean</td>
<td>7</td>
<td>32.5</td>
<td>25.5</td>
</tr>
</tbody>
</table>

**Bolded** = included in JTF list of MTI

* = labelled as a check tag
MTI can make erroneous assumptions based on populations suggested in abstract.

<table>
<thead>
<tr>
<th>Article Title</th>
<th>MTI check tags</th>
<th>Human check tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertensive emergencies</td>
<td><em>Pregnancy</em> [0]; <em>Female</em> [1]; <em>Humans</em> [2]</td>
<td><em>Humans</em></td>
</tr>
<tr>
<td>Application of the Woman Abuse Screening Tool (WAST) and WAST-short in the family practice setting</td>
<td><em>Humans</em> [0]; <em>Male</em> [1]; <em>Adult</em> [2]; <em>Middle Aged</em> [3]; <em>Female</em> [4]</td>
<td><em>Adult, Female, Humans, Middle Aged</em></td>
</tr>
<tr>
<td>A comparison of performance on the OMSITE and ABOMS written qualifying examination</td>
<td><em>Humans</em> [0]; <em>Male</em> [1]; <em>Female</em> [2]</td>
<td><em>Humans</em></td>
</tr>
</tbody>
</table>
Summary of Findings:

- In sample (N=20), more MeSH terms & accuracy were seen in the **high-JIF articles from 2000**
- High retrieval rates for human-indexed main headings & check tags; however, MTI ranking mechanisms were not consistently reliable
- Check tags reflect a certain bias for **male populations that are not aged**
- More frequent & accurate use of **medical, operationalizable MeSH terms** than **social and emotional concepts / MeSH**
Implications:

- Check tag inaccuracy is related to MTI processing abstracts rather than full texts (Mork et al., 2017)
- MTI output would benefit from greater degree of indexer review
- PubMed/Medline end users are encouraged to report problems to NLM Support Center
Limitations & Future Research:

Limitations of this research:
- Our small sample of articles in Medline does not yield strong, generalizable findings... & it therefore cannot represent all Medline articles
- The Interactive MTI tool we used is dated & may not be representative of MTIA (2022) performance

Future research / directions:
- Monitor & track indexing biases & anomalies
- Collaborate with other scholars, researchers
- Involve indexers / subject experts in projects comparing MTIA indexing to past human indexing


For a complete bibliography of our project, see [https://osf.io/4k69q/](https://osf.io/4k69q/)
Questions?

Eileen Chen, MLIS  
School of Information, UBC  
eileen.0415@livemail.tw

Dr. Julia Bullard, MA, MLIS, PhD  
School of Information, UBC  
julia.bullard@ubc.ca

Dean Giustini, MLS, MEd  
UBC Biomedical Branch Library, UBC  
dean.giustini@ubc.ca

Our project on OSF  
https://osf.io/4k69q/