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

*Eileen Chen, Dr. Julia Bullard, **Dean Giustini
*School of Information, University of British Columbia
**UBC Biomedical Branch Library, University of British Columbia

Introduction

The Medical Text Indexer (MTI) is an automated indexing program introduced in 2002 by the U.S. National Library of Medicine (NLM) to produce both semi- and fully-automated indexing recommendations based on its Medical Subject Headings (MeSH®) controlled vocabulary. As of July 2022, the NLM has moved to the most recent version of the MTI, the MTI-Auto (MTIA) (NLM, 2022), which is not well-known to the public, to index all new Medline citations.

Since NLM's release of the MTI in the early 2000s, few studies have been published on the quality of its indexing compared to human indexing. Here, we ask some important questions about NLM's planned transition to automated indexing: 1) What are the recurring errors, if any, perpetuated by the MTI? 2) What biases arise from its use of statistical and machine learning methods to retrieve and rank MeSH terms to describe the biomedical literature? 3) What limitations arise from MTI's automated indexing of abstracts rather than full-text documents?

Our aim is to compare the performance of the publicly-available version of the MTI for a set of articles indexed originally by human indexers, and to map its relative strengths and limitations.

Background

In 2007, a key user-centred evaluation of the MTI (Ruiz & Aronson, 2007) found that NLM indexers increasingly used and trusted its recommendations; however, 75% were not confident with them and 78% wanted indexing beyond a paper's abstract into the full document. More recently, its success rate depends on studies describing a clear population in the abstract (Mork et al., 2017), and on abstracts having structured sections (NLM, 2018)**.**

A majority of indexers in Ruiz and Aronson's 2007 study wanted more information about MTI's decision algorithms, and requested a better platform for indexer training. However, as NLM transitions towards automated Medline indexing in 2022, the involvement of indexers remains unclear to us. A greater reliance on automation may lead to overly standardised results that are technically correct but not contextually optimised, and human reviewers accustomed to

automated outputs may default to error identification rather than critical appraisal of context (Ehrensberger-Dow & Massey, 2014; Todorova, 2020).

The MTI's use of statistical and machine learning methods to assign MeSH terms makes it susceptible to several biases. Further, the use of the *PubMed Related Citations* feature to identify indexing terms in similar articles (and machine learning to identify subheadings), relies heavily on existing data in the indexed literature. MTI is likely to inherit or aggravate any existing human indexing problems, such as poorer indexing quality in several under-represented fields (Murphy et al., 2003; Portaluppi, 2007; Wieland & Dickersin, 2005).

This pilot study was designed to examine errors and biases resulting from these limitations.

Methods

To create our sample, we randomly selected twenty (20) articles from journals listed in the *Abridged Index Medicus* (AIM) list, a now-retired MEDLINE subset. We focused on articles published in the year 2000 – two years before the MTI was introduced – that had abstracts and MeSH terms assigned by human indexers. We selected ten articles from journals with the highest 2020 Journal Impact Factors (JIF) out of the AIM's list, and 10 from the lowest 2020 JIFs.

We entered the titles and abstracts for each article into the freely-available Interactive MTI tool (https://lhncbc.nlm.nih.gov/ii/tools/MTI.html) twice: 1) first, to generate a concise list of MeSH terms representing its final output, called the Just the Facts (JTF) list, and 2) second, to generate a full list with information on how terms were ranked by the MTI algorithm, called the Full Listing. For each article, we collated and cross-compared 14 fields of data from MTI's two lists and the human-indexed list for numbers of terms from each list, as well as the numbers of exact term matches, and the lists of identical and distinct terms, including their MTI rankings.

Results

We found that the MTI assigned more terms to articles from journals in the high JIF group than in the low JIF group, with a mean difference of 6.4 terms. Human indexers, in comparison, also assigned more terms to articles in the high JIF group, but with a lower mean difference of 2.3. On analysing three articles with the highest and lowest numbers of MeSH terms, we found the MTI ranked medical and operationalizable terms more highly. These terms were more prevalent in the high JIF group than in the low JIF group, which comprised allied health subjects such as physical therapy and nursing.

Taking the Full Listing into account, MTI had a high retrieval rate for human-indexed main headings, only missing two across 20 articles. There were also 19 instances in which the MTI used an acceptable synonym for a human-indexed term in the JTF list.

MTI consistently identified the check tag “Human” in all relevant instances, but omitted “Aged” from its JTF list in all five instances when it was used by a human indexer, leaving it out of its Full Listing 4/5 times. When a human indexer used both “Male” and “Female” check tags, MTI ranked “Male” higher than “Female,” with a ranking difference ranging from 3 to 61 places.

Conclusion

We found that the original MTI carries over some existing MeSH biases for mainstream medical topics over under-represented allied health topics. We confirmed some problems with check tag identification which may be attributed to a lack of full text access (Mork et al., 2017), and identified a bias for male populations that do not fall under the category of “Aged.”

MTI’s shortcomings may not only affect MEDLINE users by compromising MeSH integrity, but may widen existing inequities between subject areas and between populations. This study aims to raise awareness of NLM’s automated indexing project, invite the knowledge organisation community to monitor changes in MEDLINE indexing, and ask the NLM to share more information about these processes for greater transparency.

While it is important to adjust MTI’s ranking mechanisms and algorithms over time, as discussed by Demner-Fushman and Mork (2016), our findings affirm that human indexers have valuable expertise to offer during and after a transition to a reliance on automated methods. NLM has not elaborated on the extent of its human oversight of the MTI, but we argue that human curation and oversight of indexing of the biomedical literature remains critical moving forward.

Limitations and Future Directions

Our pilot study is limited by its small sample size, and our lack of access to current MTI technology. Thus, its findings may not necessarily apply to the current MTIA or to any other indexed articles in MEDLINE. Still, we will be monitoring MTI indexing issues in MEDLINE based on this pilot, and hope to use our findings to generate a more complete comparison between current MTIA quality and human indexing standards. Future studies should examine a larger, more diverse sample of papers, and involve the evaluation of indexers and subject experts.

References

To see a more complete list of references for our research, refer to our Zotero bibliography
https://www.zotero.org/deangiustini/collections/6BI9C72U
also, via the Open Science Framework: https://osf.io/4k69q/

Demner-Fushman, D., & Mork, J. (2016). *A Report to the Board of Scientific Counselors April 2016.* https://lhncbc.nlm.nih.gov/LHC-publications/PDF/pub9359.pdf

Ehrensberger-Dow, M., & Massey, G. (2014, 31). *Constraints on creativity: The case of CAT tools*.

Mork, J., Aronson, A., & Demner-Fushman, D. (2017). 12 years on—Is the NLM medical text indexer still useful and relevant? *Journal of Biomedical Semantics*, *8*(1), 8. https://doi.org/10.1186/s13326-017-0113-5

Murphy, L. S., Reinsch, S., Najm, W. I., Dickerson, V. M., Seffinger, M. A., Adams, A., & Mishra, S. I. (2003). Searching biomedical databases on complementary medicine: The use of controlled vocabulary among authors, indexers and investigators. *BMC Complementary and Alternative Medicine*, *3*, 3. https://doi.org/10.1186/1472-6882-3-3

National Library of Medicine. (2018). *Structured Abstracts* [FAQs, Help Files, Pocket Cards]. U.S. National Library of Medicine. https://www.nlm.nih.gov/bsd/policy/structured_abstracts.html

National Library of Medicine. (2022). *Frequently Asked Questions about Indexing for MEDLINE*. https://www.nlm.nih.gov/bsd/indexfaq.html#selected

Portaluppi, F. (2007). Consistency and Accuracy of the Medical Subject Headings® Thesaurus for Electronic Indexing and Retrieval of Chronobiologic References. *Chronobiology International*, *24*(6), 1213–1229. https://doi.org/10.1080/07420520701791570

Ruiz, M. E., & Aronson, A. (2007). *User-centered Evaluation of the Medical Text Indexing (MTI) System*. National Library of Medicine.

Todorova, E. T. (2020). *The impact of CAT tools on the creativity of students of Translation and Interpreting* [Newcastle University]. https://termcoord.eu/wp-content/uploads/2020/10/Todorova_BA-thesis.pdf

Wieland, S., & Dickersin, K. (2005). Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *Journal of Clinical Epidemiology*, *58*(6), 560–567. https://doi.org/10.1016/j.jclinepi.2004.11.018

Glossary of Terms

- **Abridged Index Medicus (AIM)** - a list of more than 120+ core clinical journals (mostly biomedical) available as a search subset limit or filter in PubMed up until 2020.
- **Algorithm** - a process or set of rules that are followed to guide decisions, especially by computers and machines. In indexing, it draws on previously-indexed documents.
- **Automated (and semi-automated) indexing** - the process of assigning indexing terms through automated methods to a text with no (or only partial) human intervention.
- **Check Tags** - population and/or subject characteristics mentioned in almost every article indexed in MEDLINE, such as *Human*, *Animal*, *Male*, *Female*, and *Child.*
- **Controlled vocabulary** - an established list (or database) of preferred terms in which all variants representing a concept are assembled under one controlled term for searching.
- **Human indexing** - the process of assigning index terms by trained human indexers.
- **Index Section at the NLM** - https://www.nlm.nih.gov/bsd/indexhome.html
- **Index terms in MEDLINE** - see ***MeSH terms***
- **Interactive Medical Text Indexer (Interactive MTI)** - a free MTI tool that allows users to generate indexing terms for up to 10,000 characters of text. It is based on the MTIFL, which was developed in 2011 and discontinued in 2021. https://lhncbc.nlm.nih.gov/ii/tools/MTI.html
- **Medical Subject Headings (MeSH)** - a controlled, hierarchically-organized vocabulary produced by the National Library of Medicine, used to index MEDLINE citations.
- **Medical Text Indexer (MTI)** - combines human NLM Index Section expertise and Natural Language Processing technology to curate the biomedical literature more efficiently and consistently. https://lhncbc.nlm.nih.gov/ii/tools/MTI.html
- **Medical Text Indexer - Auto (MTIA)** - the algorithm currently used for indexing is a version of the Medical Text Indexer called the sMTIA.
- **MEDLINE** - the National Library of Medicine's (NLM) premier bibliographic database of 34 million citations, freely-available online via pubmed.gov.
- **National Library of Medicine (NLM)** - the world's largest biomedical library and producer of the bibliographic database MEDLINE, available online via pubmed.gov.
- **Natural Language Processing** - a branch of artificial intelligence concerned with giving computers the ability to understand text in the same way human beings understand it.
- **Pubmed.gov** - a freely-available NLM website, containing MEDLINE and other sources, giving end-users the ability to search biomedical literature from the 1940s to present.