The NLM Medical Text Indexer System for Indexing Biomedical Literature

James G. Mork[1], Antonio J. Jimeno Yepes[2,1], Alan R. Aronson[1]

[1]National Library of Medicine, Bethesda, MD, USA
{mork,alan}@nlm.nih.gov
[2]NICTA Victoria Research Lab, Melbourne, Australia
antonio.jimeno@gmail.com

Abstract. In the face of a growing workload and dwindling resources, the US National Library of Medicine (NLM) created the Indexing Initiative project in the mid-1990s. This cross-library team's mission is to explore indexing methodologies that can help ensure that MEDLINE and other NLM document collections maintain their quality and currency and thereby contribute to NLM's mission of maintaining quality access to the biomedical literature. The NLM Medical Text Indexer (MTI) is the main product of this project and has been providing indexing recommendations based on the Medical Subject Headings (MeSH) vocabulary since 2002. In 2011, NLM expanded MTI's role by designating it as the first-line indexer (MTIFL) for a few journals; today the MTIFL workflow includes about 100 journals and continues to increase. Due to a close collaboration with the Index Section at NLM, MTI continues to grow and expand its ability to provide assistance to the indexers. This paper provides an overview of MTI's functionality, performance, and its evolution over the years.

Keywords: Indexing methods, Text categorization, MeSH, MEDLINE

1 Introduction

The NLM Medical Text Indexer (MTI) system [1] is the primary product and focus of the Indexing Initiative [2]. MTI produces both semi- and fully-automated indexing recommendations based on the Medical Subject Headings (MeSH®)[1] controlled vocabulary and has been in use at NLM since 2002. MTI is in daily use to assist Indexers, Catalogers, and NLM's History of Medicine Division (HMD) in their indexing efforts. Every weeknight MTI provides recommendations for approximately 4,000 new citations for Indexing and processes a mixed file of approximately 7,000 old and new records for both Cataloging and HMD. MTI was also used on a regular basis between 2002 and 2012 to provide fully-automated keyword indexing for NLM's Gateway[2] meeting abstract collection, which was not manually indexed. In 2011, MTI was designated as the First-Line Indexer (MTIFL) for 14 journals (89 in 2013)

[1] http://www.nlm nih.gov/pubs/factsheets/mesh html
[2] http://www.nlm nih.gov/pubs/factsheets/gateway html

because of its success with those publications. For MTIFL journals, MTI indexing is treated like human indexing and, of course, subject to the normal manual review process. MEDLINE® Indexers and Revisers consult MTI recommendations for approximately 58% of the articles they index, and the MTI recommendations are tightly integrated into the Cataloging and HMD system. Although mainly used in indexing efforts for processing MEDLINE citations[3] consisting of identifier, title, and abstract, MTI is also capable of processing arbitrary biomedical text. MTI provides an ordered list of MeSH Main Headings (MH), Subheadings (SH), and CheckTags (CT)[4] as a final result. MHs are the main descriptors or headings from the MeSH Vocabulary (e.g., *Lung*). SHs are used to qualify the MHs (e.g., *Lung/abnormalities* means that the article is about the *abnormalities* associated with the *Lung* more than the *Lung* itself), and CTs are a special type of MHs that are required to be included for each article and cover species, sex, human age groups, historical periods, pregnancy, and various types of research support (e.g., *Male*).

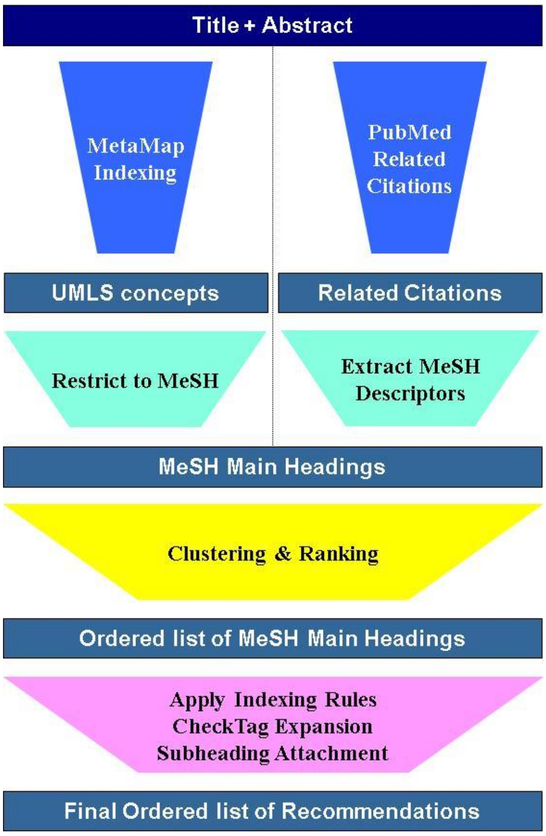

Fig. 1. MTI Process Flow Diagram

2 Processing Overview

The Indexing Initiative explored several indexing methods [2] eventually implementing two of the best ones as a prototype indexing system which became the NLM Medical Text Indexer (MTI). Normal MTI processing involves receiving a daily XML formatted MEDLINE[5] file which contains a list of Completed, In-Process, and In-Data-Review citations and a list of Deleted PMIDs (PubMed® Unique Identifier). All processing is done offline, and the MTI results are then stored in a database for later use by the Indexers. This preloading of the results is necessary since MTI takes too long to be done in real time for the Index-

[3] http://www.nlm nih.gov/bsd/mms/medlineelements.html
[4] http://www.nlm nih.gov/mesh/features2003.html
[5] http://www.nlm nih.gov/bsd/licensee/elements_descriptions.html

ers. Fig. 1 depicts the processing flow as MEDLINE citations are processed through the various components of the MTI system. Each of the major MTI components is described briefly below.

MetaMap Indexing (MMI) [3]: a method that applies a ranking function to concepts found by MetaMap [4]. Generally speaking, the MMI ranking function was designed to indicate the characterizing power or "aboutness" of a given concept for a piece of text, e.g., a MEDLINE citation. It is the product of a frequency factor and a relevance factor, which is essentially measured by MeSH Tree depth. For concepts found in the title of the citation, there is a simplified form of the function which maximizes the frequency factor.

PubMed Related Citations [5]: the neighbors of a document are those documents in the database that are the most similar to it. The similarity between documents is measured by the words they have in common, with some adjustment for document lengths. MTI currently uses two methods for determining PubMed Related Citations (PRC) for the text it is processing. If MTI is working with a MEDLINE citation and there are enough indexed PRC defined by the PubMed system[6], MTI uses that list of PRC. If MTI is processing free form text or there is an insufficient number of indexed PRC, MTI will default to using the in-house TexTool[7] implementation of PRC. MEDLINE is the indexed subset of PubMed.

Restrict to MeSH [6]: a method which finds the closest MHs to UMLS® Metathesaurus®[8] concepts. Three basic approaches can be used to map a UMLS concept to MeSH: through synonyms, through built-in mappings, and through inter-concept relationships. These approaches can be combined into a strategy that maximizes both specificity (selected MeSH terms are relevant) and sensitivity (the number of concepts that fail to be mapped to MeSH is small).

Extract MeSH Descriptors: retrieving the MeSH Heading lines from the PRC in MEDLINE format and tracking whether the MeSH Heading is a main (starred) term or not. Note that MTI does not recommend main vs. non-main status to the Indexers, but the status is tracked internally to see if MTI is improving or not.

Clustering and Ranking [7]: the ranked lists of MHs produced by the methods described so far must be clustered into a single, final list of recommended indexing terms. The task here is to provide a weighting of the confidence or strength of belief in the assignment, and rank the suggested headings appropriately.

Post-Processing: once all of the recommendations are ranked and selected, validation of the recommendations is done based on the targeted end-user. Typically, CTs are

---

[6] http://www.nlm nih.gov/pubs/factsheets/pubmed html
[7] http://www.ncbi nlm nih.gov/CBBresearch/Wilbur/IRET/TexTool/
[8] http://www.nlm nih.gov/pubs/factsheets/umlsmeta.html

added based on triggers from the text and for the remaining recommended headings, a machine learning algorithm is applied adding frequently occurring CTs [8,9], and then finally MTI performs subheading attachment [10-12] to individual headings and for the text in general.

Not all citations processed by MTI go through all of the components listed above. MTI has various filtering levels and special handling rules which require different processing pathways. Basic filtering rules have evolved over time based on ambiguities in the UMLS Metathesaurus, ambiguity in the text, feedback from Indexers, etc. Section 3 describes some of these basic filtering rules, different pre-defined levels of filtering, and some of the special handling that is required of citations.

3 MTI Filtering and Post-Processing

MTI has three levels of filtering which can be selected depending on the circumstances. *Base Filtering*, or *High Recall Filtering*, is performed for all citations and free text, regardless of whether any further filtering has been selected or not. *High Recall Filtering* is used for MEDLINE indexing recommendations and tends to provide a list of approximately 25 recommendations with most of the good recommendations near the top of the list. *Balanced Recall/Precision Filtering* provides filtering which looks at the compatibility and context of the recommendations based on what path(s) made the recommendation and provides a good balance between number of recommendations and the filtering out of good recommendations. *Balanced Recall/Precision Filtering* was developed for use in the fully-automatic processing of the NLM Gateway abstracts and is now used for MTIFL processing (see Section 5 for details). *High Precision Filtering* is the last filtering option and provides the highest level of accuracy by requiring recommendations to come from both MetaMap (MMI) and PubMed Related Citations (PRC). This provides a small list of quality MTI recommendations while filtering out many good recommendations as well. The *High Precision Filtering* option is not currently used since it provides such a short list of recommendations. Each of these filtering levels is now described in more detail.

3.1 High Recall Filtering

High Recall Filtering is designed to provide recommendations biased more towards Recall than Precision. The Indexers use the MTI recommendations as a “pick list” where they simply select the appropriate recommendations to include, thereby speeding up the indexing process. This approach tolerates some incorrect recommendations, but the majority of the recommendations need to be accurate. Recent discussions have moved MTI towards a more balanced approach where a smaller list of recommendations with a higher Precision is provided, but the list is still slightly biased towards Recall.

Terms recommended by both the MetaMap (MMI) and PubMed Related Citations (PRC) paths are subjected to a simple triage designed to immediately remove known troublesome terms. For example, all CheckTags (CT) are removed from the PRC previously indexed terms because most would be unsuitable for the article being processed. The CTs are also removed so that the recommended CTs reflect only the final validated list of MTI recommendations. Similarly, all MMI terms generated by any acronym/abbreviation of three characters or less are removed because they were triggering incorrect MeSH Geographical recommendations (for example, *T* triggered 'Texas' because a variant of *T* was *TX*). MTI also uses a hand-curated list of special cases to remove terms from the MMI path due to unfortunate variants, brand names consisting of common words, or ambiguity. For example, *sealed* in the text would trigger the MH 'Seals, Earless' because *seal* is a lexical variant of *sealed*.

The scores of certain types of terms receive additional boosting. At the beginning of each new MeSH Indexing year (usually mid-November), all of the new MH are given a special boosting by MTI that forces them to be recommended regardless of score. This is done for two reasons: 1) since they are new MHs, there will be no history in the PubMed Related Citations which would cause an artificial handicapping of the scores, and 2) to help the Indexers who might not be as familiar with the new MHs. If a MH is identified as occurring in the title of the citation, its score is tripled because terms found in the title tend to be more important. The final boosting rule floats chemicals so they appear higher up the list of recommendations and appear next to their Heading Mapped To (HM) to make identification for the indexer easier.

Next, substitution of MeSH Subheadings (SH) for certain MHs from a lookup list is done. For example, if MTI were going to recommend the MH 'General Surgery', it will be changed to the SH 'surgery'. This substitution is done because it follows the standard indexing policy where the indexer would use the SH 'surgery' in this case to qualify the purpose of the surgery. So, surgery ('General Surgery') for breast cancer ('Breast Neoplasms') becomes 'Breast Neoplasms/surgery' in the indexing.

A review of the surviving MTI recommendations is done where all recommendations that came only from the PRC path with fewer than four of the top 10 related articles providing the term are removed. It was noticed that many of the PRC path terms that were incorrect and unrelated to the text being processed by MTI had fewer than four related articles.

Finally, the list is resorted based on the changes made to the scores during filtering.

3.2 Balanced Recall/Precision Filtering

Balanced Recall/Precision Filtering was designed to mediate between the two main paths, MMI and PRC, used in MTI. MMI tends to provide more general terms, while PRC provides more specific terms which are occasionally completely unrelated to the text being processed due to normal variation in related citations. A set of heuristics was developed [7] to balance the results from both MMI and PRC by using the context of the terms they each provide. For example, one of the heuristics is "Remove any term coming from only the MMI path if either MMI or PRC provides a more specific term." This heuristic uses the context of the provided terms and the hierarchy in the MeSH Vocabulary Tree to remove more general terms typically provided by MMI. A second heuristic is "Remove any term coming only from the PRC path if MMI has not provided a more general term." Again, this uses the context and MeSH tree structure to identify PRC terms that are probably unrelated to the text. By comparing terms provided by the two paths, Medium Filtering provides a much smaller list that is more accurate (higher Precision), but still contains a reasonable number of accurate terms (acceptable Recall).

3.3 High Precision Filtering

High Precision Filtering is the simplest filtering approach - it removes any recommendation that did not come from both the MMI and PRC paths. This creates a small list of very accurate recommendations but tends to remove many good recommendations along with the bad ones. In some cases no recommendations can be made.

3.4 Post-Processing

Once filtering is accomplished, post-processing is performed regardless of the filtering level used. Post-processing involves cleaning up the final recommendation list by removing any terms that survived the filtering process but are invalid for the target audience, filling out the list of terms by adding CTs, Geographicals, and other MHs based on the text, a machine learning algorithm, and lookup lists, and then finally attaching subheadings to the individual MHs and creating a global list of subheadings applicable to the text.

The first post-processing step involves identifying the end user so the correct exclusion list can be used to remove terms from the recommendation list. There are three distinct exclusion lists used by MTI to provide tailored results for Indexing, Cataloging, and HMD. For example, the MH 'Academic Dissertations' is not used by Indexing or Cataloging, but is needed for HMD. The Indexing exclusion list is the default used by MTI and contains MHs that are too general to be recommended or contain "not used for indexing" in the Annotation field of its MeSH record (e.g., the general MH 'Eye Manifestations' with treecode C11.300 in 2013 MeSH).

The tailored recommendation list and text is then reviewed: CTs, Geographical MHs, and other MHs and SHs are added and marked so that they can be displayed as final recommendations. For example, if the MH 'Neonatal Screening' is being recommended, MTI automatically adds CTs 'Humans' and 'Infant, Newborn' if they are not already in the list. If the text contains the word *Nairobi*, the Geographical MH 'Kenya' is added to the list if it not already there. A secondary check is done for *Nairobi* to make sure the text is actually about the country *Kenya* since there is also the possibility that the text is referring to 'Nairobi Sheep'. MTI has a small set of cases like this which require a secondary check before the MH is actually added to the final recommendation list.

One final class of additions is a "forced list" of triggers whose presence within the text triggers one or more MHs. The "forced list" comes mainly from Indexer Feedback that indicated "if you see xyz, you should always recommend 'abc'." For example, if *hiv patient* is in the text being processed, MTI will always recommend the MH 'Acquired Immunodeficiency Syndrome'. MTI performs a case-insensitive search of the text for the "forced list" triggers and then adds the MH(s) if not already present and sets the "forced" flag that tells MTI to always display the term.

3.5 Subheading Attachment

MTI's final step in creating its indexing recommendations is to perform subheading attachment [10-12]. Subheading attachment is currently only done for the Indexers since Cataloging and HMD do not utilize subheadings. Due to the complexity of the data manipulation required for subheading attachment, it is not provided as a user option to MTI. Subheadings are not attached to every MH recommended by MTI; the subheading attachment algorithms use several linguistic and statistical methods to determine what is appropriate for each MH based on the text and which subheadings are allowable for each MH. MeSH specifies a subset of the subheadings that are allowed for each MH, so the subheading attachment algorithms utilize these rules to ensure that non-allowed combinations are not recommended by MTI. Based on the results of two user-centered studies [13,14], at most three subheadings are attached to each MH.

4 Improving MTI Performance using Machine Learning

Since MeSH indexing can be viewed as a categorization task, we use machine learning in an effort to improve both Recall and Precision on the most frequently used terms in MeSH [8,9]. There are some problems to consider when applying machine learning to MeSH indexing [15,16]:

- An imbalance between positive and negative instances,
- Even if a MeSH heading is correctly identified with a citation it might not be significant enough to be included in the indexing,
- There are inconsistencies between indexers, and
- Changes are made to indexing policy over time.

From the set of MHs, we selected the 40 most frequently indexed MHs. Most of these ended up being CTs or MHs that MTI treats like a CT (e.g., 'Swine'). We compared several learning algorithms that we could run efficiently on a training set of 200k citations, and selected the best learning algorithms for each one of the MHs based on a meta-learning approach. The results of various experiments with machine learning provided improvements for twelve of the MHs identified in Table 1 compared to MTI [9]. The table shows the CT, MTI F_1 scores prior to and after implementing the machine learning algorithms, and how much of an improvement is obtained for each CT. The Precision and Recall for calculating the F_1 score are based on comparing the human indexing as the gold standard against the MTI recommendations for each citation. The trained machine learning algorithms for these twelve MHs were incorporated into the MTI processing flow. We used our MTI_ML machine learning package[9] in these experiments. The MTI_ML package was developed as part of the Indexing Initiative effort to provide machine learning algorithms optimized for large text categorization tasks and capable of combining several text categorization solutions. It is available subject to the MetaMap terms and conditions[10].

Table 1. CheckTags Before and After Machine Learning Applied

CheckTag	F_1 prior to ML	F_1 with ML	Improvement
Adolescent	0.2475	0.4236	+0.1761
Adult	0.1949	0.5684	+0.3735
Aged	0.1172	0.5467	+0.4295
Aged, 80 and	0.0150	0.3089	+0.2939
Child, Preschool	0.0611	0.4540	+0.3929
Female	0.4606	0.7384	+0.2778
Humans	0.7998	0.9133	+0.1135
Infant	0.3439	0.4469	+0.1030
Male	0.3847	0.7114	+0.3267
Middle Aged	0.0101	0.5950	+0.5849
Swine	0.7104	0.7475	+0.0371
Young Adult	0.0283	0.3163	+0.2880

At indexing time, the text of the citation is provided to the trained learning algorithms and a result for each of the above twelve MHs is provided stating whether to add the term to the list of MTI recommendations. Additions are added as “forced” terms meaning that they are guaranteed to be in the final MTI recommendation list.

We have been working to improve these results by adding learning algorithms to our training step, like SVM based on Hinge loss and modified Huber loss and we have evaluated the performance of combining several learning algorithms [9]. The experiments have considered not only the most frequent MHs, but a selection of MHs using several criteria. As future work, we would like to evaluate more learning algorithms,

[9] http://ii nlm nih.gov/MTI_ML/index.shtml
[10] http://metamap.nlm.nih.gov/MMTnCs.shtml

e.g. additional kernels, experiment with several feature engineering approaches, and see if machine learning might improve some of the poor performing MHs.

5 MTI First-Line Indexer (MTIFL)

In 2010, the Indexing Initiative team and the Index Section conducted a series of three experiments with MTI to determine the feasibility of using MTI recommended MHs as first-line indexing for selected subject areas. Journals for the three experiments were chosen from fields where MTI was performing well (for example, *Microbiology*, *Anatomy*, *Botany*, and *Medical Informatics*). The experiments measured the accuracy of MTI indexing and the amount of time required to index and revise both the manual and MTIFL indexing. The results of the experiments showed that MTIFL was successful given the right circumstances, namely journals with a low potential for the need of manually created chemical flags and GeneRIFs that are normally added by the indexer. In the case of MTIFL, the burden of creating the chemical flags and GeneRIFs would shift to the reviser which would be time consuming and undesirable.

MTIFL partially automates the standard indexing process, which consists of two steps: 1) indexers assign MeSH to describe the content of an article based on a review of the full text, and 2) in-house revisers, senior staff who are expert indexers, review and modify the indexing and release it for searching and viewing in PubMed. MTIFL uses MTI for the first step of indexing, focusing on only the titles and abstracts. In-house revisers continue with the second step, reviewing the MTIFL indexing, adding or deleting MHs, and releasing the final indexing to PubMed.

In February 2011, fourteen journals were initially selected to be included in the MTIFL pilot, and 75 journals have been added since for a total of 89. The process of evaluating additional journals for inclusion in the MTIFL program is ongoing. One outcome of the MTIFL experiments was that the timing information showed it took indexers longer to remove wrong MTI recommendations than to add missing ones. In talking with the indexers, the reason for this extra time when removing a bad recommendation is that they have to take time and decide if they missed something in the article or not before removing it. So, MTIFL journals are processed with MTI's Balanced Recall/Precision Filtering option providing a smaller, more precise indexing list than with the regular processing. The average F_1 measure increases by 0.1889 when journals are incorporated into the MTIFL program. This increase is likely due to the extra filtering and indexing policies specific to MTIFL.

6 MTI Performance

MTI has shown a steady increase in usage and acceptance by the NLM indexers since 2002 when it first started producing recommendations for them. MTI is now a mature indexing tool that benefits greatly from a close collaborative relationship with its customers. The strides that MTI has been able to make over the last two years would not be possible without the continued collaboration with the Index Section providing much needed expertise and insight to the indexing task.

We use the human indexing as a gold standard and compare that against the MTI recommendations to calculate Precision and Recall. Table 2 displays Precision, Recall, and F_1 measure for Overall MTI performance from 2008 to 2012 the last full year for which we have statistics. The last column shows the differences between 2008 and 2012. The table also shows how, over the years, MTI has changed and improved. It clearly shows that between 2008 and 2012 there is a shift towards more precise recommendations with increases across the board in Precision statistics and only very slight gains in the Recall. The table also shows that MTI was able to provide recommendations for over 93% of the total number of citations that were indexed in 2012.

Table 2. MTI Performance though the Years

Overall	2008	2009	2010	2011	2012	Diff
Recall	0.5258	0.5381	**0.5127**	**0.4740**	**0.5554**	+0.0296
Precision	0.3068	0.3103	**0.3268**	**0.5157**	**0.5410**	+0.2342
F_1	0.3875	0.3936	0.3992	0.4940	0.5481	+0.1606
Citations with MTI Recommendations	606,566	684,599	664,905	694,552	711,863	+105,297
Citations Indexed[11]	671,904	712,675	699,420	724,831	760,903	+88,999
% MTI Recommended	90.28%	96.06%	95.07%	95.82%	93.56%	+3.28%

Taking a closer look at the Recall, Precision, and F_1 measures for just 2010 – 2012, the dramatic changes are easier to see in the performance graph shown in Fig. 2. The big changes at the beginning of 2011 show the effects of MTI's change in filtering to focus on Precision over Recall based on the results of our MTIFL experiment findings. Precision rose from 0.3268 to 0.5157 (+0.1889) during this time and Recall fell from 0.5127 to 0.4740 (-0.0387). The further increases for both Precision and Recall at the beginning of 2012 were due to improvements in MTI rules and the inclusion of the machine learning algorithms for the 12 CheckTags listed in Table 1. The fact that we were able to improve both Precision and Recall is credited to the machine learning algorithms providing us a much needed boost in Recall along with excellent Precision on the 12 frequently occurring CheckTags. Recall improved from 0.4740 to 0.5554 (+0.0814) and Precision improved from 0.5157 to 0.5410 (+0.0253).

[11] http://www.nlm.nih.gov/bsd/bsd_key html

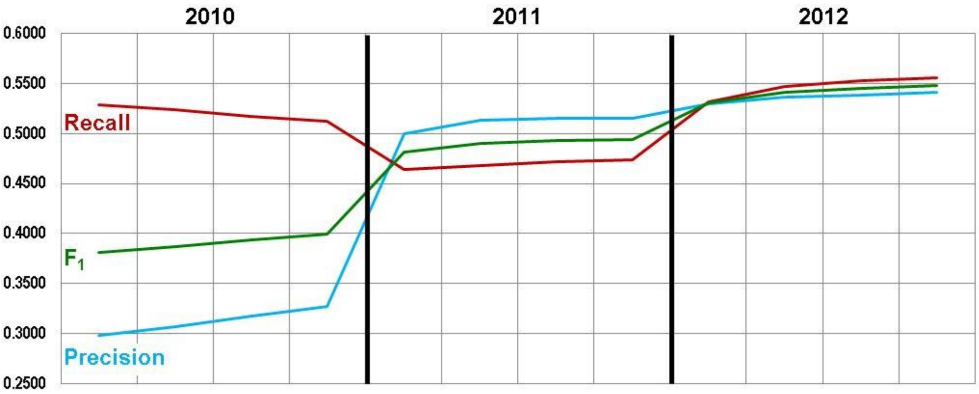

Fig. 2. Closer look at MTI Performance 2010 - 2012

Table 3 shows Precision, Recall, and F_1 measure for MTIFL performance from 2011 to 2012. MTIFL has greatly expanded the assistance that MTI can provide and has increased the pressure on MTI to continually improve. In Table 3 we can see a dramatic increase in the F_1 measure for MTIFL journals over regular MTI (0.5481 MTI from Table 2 versus 0.7152 MTIFL from Table 3), and care needs to be taken to make sure that these increases are due to actual MTI improvements and not to changes in indexer behavior. Indexers are told to accept MTIFL indexing that is not incorrect and to correct only that which is wrong -- meaning that MTIFL indexing is treated the same as a human indexer. This differs greatly from the normal indexing process where MTI is simply used as a tool for indexers to use or not use as they wish. So, the enthusiasm for the dramatic increases has to be tempered with the knowledge that some of the change is due to how MTIFL is used and not to improvements in the program itself.

Table 3. MTIFL Performance

MTIFL	2011	2012
Recall	0.6111	0.6964
Precision	0.6780	0.7351
F_1	0.6428	0.7152
Citations	3,435	4,205

Future Direction

The Medical Text Indexer Team benefits from a very close collaboration with the NLM Index Section. This collaboration provides a deeper understanding of the manual indexing process and insights into other possible avenues where MTI might be used to assist in the indexing process at NLM.

Several research topics that are planned for the future include: utilizing full text now that it is becoming more available, assisting in Gene Link and Chemical Flag identification, utilizing sections identified in Structured Abstracts to help weight recommendations, identify whether author/publisher supplied keywords might benefit MTI, and expanding machine learning usage to help improve problematic MeSH Headings. We also look forward to expanding the number of MTIFL journals.

Acknowledgements

This work was partly supported by the Intramural Research Program of the NIH, National Library of Medicine. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. We would like to thank our colleagues François Lang and Willie Rogers for providing direct and indirect support of MTI. We would also like to extend special acknowledgment to Hua Florence Chang who was the original creator of MTI. Florence's foresight has provided us with a robust and tunable program.

References

1. Aronson AR, Mork JG, Gay CW, Humphrey SM, Rogers WJ. The NLM Indexing Initiative's Medical Text Indexer. Medinfo. 2004 Sept.;2004: 268-272.
2. Aronson AR, Bodenreider O, Chang HF, Humphrey SM, Mork JG, Nelson SJ, Rindflesch TC, and Wilbur WJ. The NLM Indexing Initiative. Proc AMIA Symp 2000;:17-21.
3. Aronson AR. The MMI Ranking Function Whitepaper (1997). Available at http://skr nlm.nih.gov/papers/references/ranking.pdf.
4. Aronson AR and Lang FM. (2010). An Overview of MetaMap: Historical Perspective and Recent Advances. J Am Med Inform Assoc. 2010 May 1;17(3):229-36.
5. Lin, J., & Wilbur, W. J. (2007). PubMed related articles: a probabilistic topic-based model for content similarity. BMC bioinformatics, 8(1), 423.
6. Bodenreider O, Nelson SJ, Hole WT, and Chang HF. Beyond Synonymy: Exploiting the UMLS Semantics in Mapping Vocabularies. Proc AMIA Symp 1998;:815-9.
7. Medical Text Indexer (MTI) Processing Flow Whitepaper. Available at http://skr nlm.nih.gov/resource/Medical_Text_Indexer_Processing_Flow.pdf.
8. Jimeno-Yepes, A., Mork, J.G., Demner-Fushman, D., and Aronson, A.R. Automatic algorithm selection for MeSH Heading indexing based on meta-learning. International Symposium on Languages in Biology and Medicine, Singapore, December, 2011.
9. Jimeno-Yepes, Antonio, Mork JG, Demner-Fushman D, Aronson AR. Comparison and combination of several MeSH indexing approaches. AMIA Annual Symposium Proceedings. Vol. 2013. American Medical Informatics Association, 2013.
10. Névéol A., Mork J.G., Aronson A.R.. Automatic Indexing of Specialized Documents: Using Generic vs. Domain-Specific Document Representations. Proc BioNLP 2007 Workshop, 183-92.

11. Névéol A., Shooshan S.E., Humphrey S.M., Rindflesch T.C. and Aronson A.R. Multiple Approaches to Fine-Grained Indexing of the Biomedical Literature. Proc Pacific Symposium on Biocomputing 2007, 292-303.
12. Névéol A, Shooshan SE, Mork JG, Aronson AR. Fine-Grained Indexing of the Biomedical Literature: MeSH Subheading Attachment for a MEDLINE Indexing Tool . AMIA Annu Symp Proc. 2007;:553-7.
13. A MEDLINE Indexing Experiment Using Terms Suggested by MTI Whitepaper, June 2002. Available at http://ii nlm nih.gov/resources/ResultsEvaluationReport.pdf.
14. Ruiz M.E. and Aronson A.R. User-centered Evaluation of the MTI System, 2007 Whitepaper. Available at http://ii nlm nih.gov/resources/MTIEvaluation-Final.pdf.
15. Jimeno-Yepes, A., Wilkowski, B., Mork, J.G., Demner-Fushman, D., and Aronson, A.R. (2012). MeSH indexing: machine learning and lessons learned. ACM SIGHIT International Health Informatics Symposium, Miami, FL, USA, 2012.
16. Jimeno-Yepes A, Mork J, Demner-Fushman D, Aronson AR. A one-size-fits-all indexing method does not exist: automatic selection based on meta-learning. JCSE, vol. 6, no. 2, pp.151-160, 2012.