

## Transcript

MIKE: Thank you all for joining us today for another NLM Office Hours. My name is Mike Davidson. I'm in the Training and Workforce Development Team which is part of the User Services and Collections Division here at the National Library of Medicine. My pronouns are he, him, his. The goal behind these Office Hours sessions is to give you a chance to learn more about NLM's products and to get your questions answered by our trainers and by members of our product teams.

Today our focus is on PubMed and we've got a great roster of folks with us today who know a lot about PubMed and a lot about different aspects of PubMed. Before we get to your questions, we're going to kick things off with a brief presentation from Alex Sticco, who is Team Lead for Automation and Strategic Analysis in our Controlled Vocabulary Services program. Alex's team are heavily involved with improvement and updates to our MEDLINE indexing process, and Alex will be sharing some information about those updates. We'll then use the rest of the session to have our panelists answer your questions about MEDLINE indexing or anything else having to do with PubMed.

A few quick logistical notes before I hand things over to Alex. We are recording today's session to share with those who are unable to attend. The recording will be posted shortly following the office hours for you to review or share with others. We have a pretty substantial crowd with us today, so we've muted all attendees. However we encourage you to submit your questions as you think of them throughout the session using the Zoom Q&A feature, you can find that at the bottom of your Zoom window. There's a little Q&A button with two word bubbles. Click on that to open that up. When we get to the Question & Answer portion of the session, I will look at that Q&A panel and direct your the questions you submitted to the right panelists so that they can answer verbally. We may also be occasionally using the chat feature to share some links to helpful resources. But again we ask you to post your questions in the Q&A tool. That way we can make sure we see them and get them to the right person. But before we get into addressing your questions, I'm going to hand things over to Alex to bring us up to speed on what's new with MEDLINE indexing. Alex, take it away.

ALEX: Great. Thank you for that introduction, Mike, and thank you everyone for joining us today for this update on MEDLINE indexing where I will be introducing our new, upgraded indexing algorithm MTIX.

So today I'm going to start by providing a little context for automation and indexing, briefly describing what the overall MEDLINE indexing pipeline is, what's changing and what is not changing. And then I will describe the basic technology behind the new algorithm and how it differs from the old one. And then finally, I will provide some details on its performance and the ways that we evaluate that performance. We want to leave plenty of time for your questions, so I'm going to strive to be brief and give this information in pretty broad strokes. If you would like more detail about anything, that's what the Q&A is for.

So let's get started with a little background on automated indexing. NLM has been using some automated indexing in MEDLINE for many years, but a transition to full automation was completed just two years ago, in April of 2022. For many years prior to automating, the volume of published literature had outstripped our ability to index it in a timely fashion. And so automation allowed us to finally eliminate a backlog that had peaked at over 850,000 citations. At the time, the lag time between when an article first appeared in PubMed and when it received indexing was typically between one and three months. And now it is just a single day. But now two years later, we are upgrading the algorithm that

performs the indexing. So the goal of this talk is to introduce you to that new algorithm which we will be deploying in the middle of March. It is not currently deployed.

So a quick point of terminology. Our current indexing algorithm and the new one have very similar names which can be a little confusing. So the current algorithm is MTIA and the new algorithm is MTIX. And you can remember A is for the beginning of the alphabet, so it is also for the beginning of automated indexing.

To start, here is the basic pipeline for indexing MEDLINE citations first. New citations are uploaded to PubMed every day, and then MEDLINE citations are indexed by MTIA within 24 hours. After that, finally, a subset of articles undergoes human curation for quality assurance, and this is generally completed within the next two weeks. So the only thing that is changing here is the indexing step which will soon be performed by MTIX instead of MTIA.

There are some things that we are not changing that we are frequently asked about, so I wanted to address those upfront. First, the scope of indexing will not change at this time. Indexing will continue to be applied only to MEDLINE journals. MEDLINE makes up the bulk of PubMed citations, but PubMed does also contain non-MEDLINE citations which we do not index and that includes preprints. At this time, there are no plans to extend indexing to those other citations. Also, indexing will continue to be based on the title and abstract and not the full text. Although we are very interested in using the full text, our licensing agreements with publishers still do not allow us computational access for what they consider data mining applications. Additionally, many full text papers are too long for the new technology to handle as input. We still hope to use full text in the future, and we will certainly keep our users updated if our process changes.

Now let's take a look at the MTIX technology and how it differs from MTIA. So these two technologies have very similar names, but they are using very different technologies. MTIA is a complex system, but it's primarily based on a dictionary of MeSH terms, synonyms, and other trigger phrases. It looks for those words in the title and abstract and assigns MeSH terms based on the frequency and the position of those matching words. So those matches are then refined with hundreds of different rules. For example, the terms patient and patients usually trigger indexing the term humans. However, we have a rule that blocks that trigger in veterinary journals where patients are more often animals. All of those rules have been created and refined by people over the course of many years, mostly based on feedback from indexers. In contrast, MTIX is a machine learning model known as a neural network. Neural networks underlie many now familiar AI applications, and MTIX is a kind of low-level AI. With this type of algorithm, instead of giving it indexing rules that we have created, we give it data to learn from and it creates its own rules. MTIX is trained on MEDLINE citations and their indexing. Based on that data, it infers relationships between the indexing on the citation and features in the text, like vocabulary, the distance between words, parts of speech, and sentence structure. When it sees a new citation, it can then use the knowledge it has developed about words and their relationships to determine which terms are statistically most likely to be the appropriate indexing. Essentially, MTIA is all heuristics. It's a big collection of rules of thumb. MTIX is all statistics. It's calculating the probability that an indexing term matches a citation.

So what difference does that make, and how these two systems perform? Well, the big one is that MTIX significantly outperforms MTIA, and a lot of that performance improvement is because MTIX is not limited to literal word matches like MTIA. It understands more complex representations of concepts like

interrupted phrases, and it can even recognize abstract ideas that may not literally be named in the text. In fact, because it is making determinations based on many features and not just trigger words, it can often predict when a MeSH concept will be present in the full text of an article, even though it is not mentioned in the abstract at all. Additionally, MTIX does not make the same kind of contextual errors that MTIA makes when encountering metaphorical language. With a more sophisticated understanding of context, MTIX knows that the elephant in the room is not *elephas maximus* and that Quaker parrots are not Protestants.

Finally, there is the issue of new MeSH terms. When new terms are added to the workflow, to the vocabulary, MTIA is able to handle them fairly gracefully. So they're simply added to the lookup dictionary, and then any relevant existing rules are applied. So, for example, rules that apply to all protein terms are applied to new protein terms as well. How good MTIA is at indexing those new terms depends on exactly the same factors that affect the old terms: how comprehensive the list of synonyms is and whether those exact terms appear in the citation text. Updating the vocabulary of MTIX is somewhat more difficult. The Achilles heel of all machine learning applications is that they can't learn new things without substantial new training data. To learn how to use new mesh terms, MTIX needs to see many examples of citations indexed with and without those new terms. Since we are no longer indexing large numbers of citations by hand, we can't supply that training data for new terminology. We are investigating various ways to generate training data for new terms. However, in the meantime, new terminology will be indexed using technology that is similar to MTIA, relying on word matching and rules that we create. So for the time being, there will not be very much difference between these two systems in the way that new terms are handled.

Finally, let's take a look at how MTIX performs. We use a number of different evaluation modes to understand the performance of our algorithms. The raw data can give a misleading impression of the actual efficacy of MTIX because some context has to be applied to interpret those numbers. So before I present the actual numbers, I want to explain the metrics we use and their limitations.

Our main evaluation method is comparing MTIX output to manual indexing. With this method, we assume that the manual indexing is the correct indexing and then assess how much MTIX gets right by comparison. The metrics we use to describe that are called recall, precision and the F1 measure. And the idea behind precision and recall is that MTIX can make two kinds of errors. First, it can miss terms that were in the original indexing. Recall addresses these missing terms. Recall is the percentage of the original indexing that MTIX did find. It lets us know how complete the MTIX indexing is. The other kind of error is adding extra terms that shouldn't be there. Precision addresses these extra terms. Out of everything that MTIX outputs precision is the percentage that it got correct. It lets us know how much extra noise is being added. And finally the F1 measure combines precision and recall in something like an average to give us a single overall score.

This is our standard evaluation for many reasons. It's obviously directly relevant to the work. It's simple and it's fast to set up these comparisons and we have lots of pre-existing indexing so we can use test sets with tens of thousands of articles to cover all the different kinds of articles and vocabulary.

But this type of evaluation also has some very significant weaknesses. The first is that manual indexing isn't technically a gold standard, meaning it's not a test set where the answers have been validated by a panel of reviewers with high inter-rater reliability. If you ask two indexers to index the same article, you will often get different results, and there are many factors that contribute to that variability. First, the

MeSH vocabulary is quite large and has a lot of overlapping concepts. Thus it's often possible to index the same concepts with different vocabulary. One way might be better, but you wouldn't look at the other options and think that they were wrong. Also, indexers also vary in their opinions about which concepts in the article are important enough to index, with some choosing to add more secondary concepts or different secondary concepts than others. And of course, indexers make errors sometimes, like all humans. The second major weakness of this evaluation is that it treats all errors as having equal importance. Realistically, some vocabulary terms will have much more impact on searchers than others. So, for example, publication types are used by searchers much more than other terms. Additionally, some errors are more wrong than others. If you have an article about hypertrophic cardiomyopathy and you index it with just cardiomyopathy, that's a more general term than we wanted, but it's still a relevant concept that isn't adding noise to the overall search. It's not as wrong as indexing that article with an unrelated concept like cardiac arrhythmias.

All of this means that when we evaluate MTIX by comparing it to existing indexing, a lot of what we count as errors from MTIX are actually improvements or even corrections to the original indexing, best case scenario. Alternative ways to represent the same concepts as the original indexing. Concepts that are present in the article but that the original indexer felt were not important enough to index. Or, real errors, but ones with limited impact. These aren't trivial details. These are quite important considerations in an evaluation of how much we're getting right or wrong. The problem is that we have no easy way to determine which errors are which except by manually evaluating each one, and we cannot do that on the scale that we generally want to perform evaluation most of the time.

However, we have done so in a small-scale manual evaluation. For this evaluation, we took 185 recent articles and we ran MTIX on them, and then we asked indexers to evaluate the errors made by MTIX when compared to the original indexing. As I mentioned, there's two kinds of errors, extra terms and missed terms. So for any extra terms MTIX had added, indexers could rate each of those terms four ways. The term could be truly incorrect, it could be an improvement or an equivalent to something in the original indexing, or it could be one of two sort of partial credit options. The missing terms were also rated by a similar rubric. They could be truly missed concepts, concepts covered by an equal or better alternative term, or partially covered by such terms, or concepts that should not have been indexed in the first place where the original indexer made a mistake.

When we examine the errors closely this way, we find that most of the extra terms aren't really errors. Almost 2/3 of the extra terms were considered improvements or equivalents to the original indexing. Another quarter, we're at least partially correct. Indexers considered only 12% of the extra terms to be true errors. Those true errors represent just 5% of the total output of MTIX. For the terms that were missed, those numbers are somewhat reversed. Almost 2/3 of missed terms were just concepts MTIX missed. About 1/5 either should not have been indexed originally or were covered with different terms that were equivalent or better, and another 16% were partially covered by other terms. So using this data, we are able to adjust the precision and recall metrics for this test set.

For the set, the original raw precision, recall, and F1, we're all in the high 60s or low 70s. This is typical or it's actually even a little lower than the raw numbers that we see overall. However, if we correct everything that was equal or improved to a true positive, then we get a dramatic increase in precision shooting from 72 to 88%. Recall sees a very modest gain of two percentage points to 69%. And the F1, our overall score goes from 69% to 78%. If we also weight the partial errors so that they don't count as

strongly against us as the true errors, then precision jumps all the way up to 92%, a full 20 percentage points above the raw number. Recall gets one more point. The overall F1 ends up at 80% instead of 69%, 11 points higher. I think it's fair to say from these results that MTIX is producing indexing that is quite clean with very little output that isn't at least relevant to the article. The recall numbers indicate that that indexing is not as comprehensive as we might like, but keep in mind that these results also suggest that the original indexing could have been more comprehensive. In fact, we believe that what we are seeing with these evaluations is that MTIX differs from indexers in ways that are very similar to how indexers differ from each other, choosing different parts of the vocabulary to represent the same ideas, making different choices about which minor concepts are important enough to index, and a few real mistakes here and there. So please keep this in mind as we look at the raw numbers from a much larger comparison set and compare MTIX performance to MTIA.

These are the raw F1 scores for MTIA for various parts of the vocabulary on a test set of about 32,000 articles. I said before that the F1 is a type of average between precision and recall, so instead of a typical average between two numbers, it is designed to penalize values that are very far apart. MTIA is tuned to have high precision at the expense of recall because we would rather add fewer indexing terms and be more certain they're the right ones. So we do have quite a large difference between precision and recall, which means that our F1 scores get dragged to the lower end. For example, the overall precision is 74%, whereas the overall recall is just 49%. And you can see that the F1 score is much closer to the lower value than to the higher one. And that's the case for all of these categories. So with that in mind, let's compare the MTIX scores.

MTIX performs significantly better in every category. This is primarily due to large gains in recall without significant reductions in precision. Because MTIX can accurately apply concepts in many circumstances where MTIA cannot. The overall precision for MTIX is 76% and the overall Recall is 72%, but you can see from these F1s that performance is superior in Publication Types and Checktags, two categories with especially high user impact.

And we know that some of you have been eager to see more details on checktag performance specifically. So here is a very crowded chart with all of the details on all 15 checktags. If you are unfamiliar with checktags, they are a special set of MeSH headings that we use to describe the research cohort of the article rather than the article topics. So if a clinical trial includes both male and female participants, that article would have the male and female checktags, even if gender or sex aren't the topics being researched. Checktags are available as filters in the PubMed left-hand sidebar, so you can use them to narrow your results according to the makeup of the research cohort. And because of that, they are used much more often in searches than other MeSH headings, so we consider them a critical area of performance for the algorithm. You can see that performance is excellent for humans and animals, and it's very good for most of the other terms. Our low spots are newborn infants, adolescents, young adults, and aged 80 and over. And this is likely because abstracts frequently do not supply enough information to detect those, or enough information to distinguish between those ages and their more common near neighbors, infant, adult, and aged. We know that many people depend on the checktags, so we hope that this information will help you adjust your searches according to how reliable a particular checktag will be.

And we are always striving to improve. So I want to end by sharing some of our upcoming plans. A major R&D effort this year will be experimenting with ways to generate training data for new MeSH terms,

which might include using assistance from other AI. We are also planning to conduct more research on the way searchers are using indexing and MeSH in PubMed. Understanding the areas of vocabulary that have the most impact on our users will allow us to focus curation, algorithm improvement, and vocabulary development on the areas that will make the most difference. And finally, we are in general keeping a close eye on what is happening in the world of AI, especially large and small language models. This is obviously an area that is moving incredibly fast. We recognize how important it is to keep reevaluating our options and strategies when the technology we are using sometimes changes significantly within just a few months. It is certainly an exciting time to be in information science. And with that, I will hand things back to Mike to get our Question & Answer period started.

MIKE: Thank you so much, Alex. And yes, it is absolutely a very exciting time. As we said, we're now going to spend the rest of this session answering as many of your questions as possible. We have had a few folks submit questions to us ahead of time, which we greatly appreciate. And I already see some questions, many questions actually in the Q&A panel. So we're off to a good start. Please keep submitting those questions, any questions you have about indexing or anything to do with PubMed right in that Q&A box. I'll read them off to our panel and have them answer your questions verbally.

Speaking of our panel, I want to give them a very brief introduction here before I start peppering them all with your questions. In addition to Alex, our panel of NLM experts includes Melanie Huston and Deborah Whitman, also part of our Controlled Vocabulary Services program. We have Amanda Sawyer, Jessica Chan, and Marie Collins from the NCBI PubMed team. And Kate Majewski, who is my colleague from the training and workforce development team. And also Michael Tahmasian, who's helping us out with the tech side of things today. So between all these folks, we should hopefully be able to answer any PubMed questions you have.

So we're going to get right into it. And I'm going to start with a couple of different questions that sort of revolve around what goes into the automated indexing algorithm, the MTIX, what are the inputs? So Melanie, I think this first one is probably for you because in reference to what Alex was saying about the input maximum with full text being too long, **what is the maximum input length for the MTIX algorithm?**

MELANIE: Yes, thanks, Mike. The maximum input that goes into training the algorithm is 512 tokens, which is basically 512 meaningful words from the title and abstract. And that is a length that's able to cover the vast majority of the titles and abstracts to get us the results that we're getting with MTIX.

MIKE: Excellent. On a similar token, sort of about what goes into the algorithm, we have a question from Stacy saying, **regarding TIAB being used for indexing, does that include journal title, affiliations, etcetera, or only TIAB, which is I think in reference to the title and abstract field in PubMed.** So I'm going to hand that to Amanda who can maybe address that.

AMANDA: Yeah, this is a great question. So in PubMed, we have the title abstract field that does include things like author keywords. When we're talking about title and abstract inputs for MEDLINE indexing, that's not necessarily directly related to that field of PubMed. So the inputs for MEDLINE indexing include the title, the actual text of the abstract itself, the full journal name, and the publication year. So those are the MTIX inputs. And then the title abstract field in PubMed is a separate concept, related but separate.

MIKE: Excellent. Thank you for that. And again on another note of inputs for these algorithms, Stacy asks, **was the training data set for MTIX selected from human-indexed or MTIA-indexed articles?** And that would probably be for Melanie or Alex.

MELANIE: I can take that one.

ALEX: Go ahead, Melanie.

MELANIE: The training data does come from human-indexed articles. We are very careful to make sure that the training citations that we choose have been manually indexed by humans and do not have input from MTIA.

MIKE: Awesome. OK, thank you for that. Another question here from, I apologize if I'm mispronouncing your name, VHAPORPAYNTR. Oh sorry. I think this would be for Deborah. **How many MeSH terms on average are assigned to a record by MTIX?**

DEBORAH: Hi, thank you for that question. Human indexers used to apply roughly 10 to 15 terms per citation. Our current algorithm, MTIA, often suggests less, you know, maybe 7 or 8. And the new algorithm MTIX is more comparable to human indexing. So on average it's suggesting around 13 terms and that does include publication types and checktags.

MIKE: All right, Thank you so much for that. We have a couple of questions here about the evaluation process that was described by Alex and possibly further evaluation as well. Stacy asks **how were the 185 articles in the test set selected and how was the number 185 determined?** Alex, if you want to take a crack at that, you can or Melanie can also probably help out with that.

ALEX: Yeah, those were selected from a larger test set that we were already using and we selected them for overlap with that test set so that we could correlate other sort of aspects of the vocabulary representation and things like that within the smaller manual set that we were doing to the larger overall test set that we were using for the, you know, sort of bigger computational evaluation. The number was actually 200 total, but fifteen of those were articles that had been indexed automatically and we wanted to sort of have a small sample like that to verify that the indexing from MTIX, if we reindexed articles that had previously been automatically indexed that the indexing from MTIX would be superior to the indexing that MTIA had already applied. So that was why it's a strange number. I only presented the results from the ones that had been manually indexed in the first place.

MIKE: Excellent. And sort of on a related note about the evaluation, there was a question about essentially **how do we know that the metrics are correct if MTIX doesn't have access to the full text?** Melanie, I think you were pointing out in our private chat that the evaluators had access to the full text, is that correct?

MELANIE: That's correct. So MTIX has been trained on titles and abstracts. We wanted to evaluate how that was performing. We did have the indexers performing this evaluation look at the full text and you saw the results of the study.

ALEX: And I would add that the original indexing on those articles was performed from the full text. So the original manual indexing is from the full text. So we know that the original manual indexing is representing the information there.

MIKE: Excellent. Sort of continuing in this evaluation vein, I'm trying to be a little bit thematic here so we're not just jumping around all over the place. Alex, another question from Stacy, **will evaluation be ongoing, are there thresholds that would prompt NLM to take corrective action?**

ALEX: So yes, evaluation absolutely will be ongoing and we have a number of different ways that we are monitoring. And of course we have the feedback from curators. I would say that there are three ways that we identify errors. We've got feedback from our curation staff, we have feedback from the public, and we also analyze performance on test sets, like the numbers that I showed you today. Often times we're going into you know many other sort of areas looking at different branches of the vocabulary and things like that. So there are a lot of evaluations that we performed that I couldn't present today because I can only show you a little bit. But there's definitely going to be a lot of ongoing evaluation and well I would say that there are no specific sort of action thresholds for performance declining. That's really because we don't anticipate performance declining over time. So it's more that when we make a change, we do a lot of upfront verification. So a lot of the time that we've spent between automating with MTIA and releasing MTIX in March has been spent sort of validating and making sure it's not doing worse than any particular categories, that we're not just looking at these overall numbers, but that we make sure that the performance is really improving across the board or that the trade-offs are things that we feel like are worthwhile because of the greater impact of one thing on our users than versus another.

MIKE: Excellent. Yeah, there's a lot of nuance in that, sort of how we respond to this information. And sort of closing out the evaluation discussion for now, I'm sure we'll get back to it, but another question from Stacy, **are the details of this evaluation published? Will they be published? Are there any plans to publish any, anything based on this evaluation?**

ALEX: I don't think there's plans to publish, they're not published. Because we're not staff scientists, and it's very time consuming to publish. We typically don't.

MIKE: It's more operational evaluation rather than scientific evaluation.

ALEX: That is correct.

MIKE: Yeah, totally understand that. All right. I'm going to change gears just for a little bit here. And where was that question that I had? Going to go into sort of a pure PubMed question. This will be for Amanda from Claudia. **Are there any plans to actually switch or add a semantic large language model-based search option complementing the classic Boolean keyword-based search in PubMed?**

AMANDA: This is a really interesting question and we're always looking for new ideas and new ways we can improve PubMed, but we don't have any plans around this concept specifically at this time. Yeah, obviously we're always looking at things and investigating things, but nothing on the plan right now.

MIKE: All right, I have another one for you, so don't go away. I'm just trying to find it again.

AMANDA: OK.

MIKE: You folks are doing great about giving us all of these questions. Ah, here we go from Tracy. **Will MTIX impact the phrase index in PubMed in any way? For example, adding new phrases to index when MTIX sees a statistical increase in those phrases. So first of all, what is the phrase index and how might this be impacted?**



AMANDA: Sure, this is a great question. Let me go ahead and share my screen while we're talking, because the phrase index is fun, a PubMed specific thing to talk about. All right, I'm sharing my screen.

This is PubMed's advanced search page. PubMed uses what we call a phrase index when you're searching for an exact phrase rather than searching for something that maybe you want to let PubMed translate to include synonyms. And we do this because this is the most efficient way we're able to provide phrase searching in PubMed. For context, remember, PubMed is more than 36 million records. We have around 3.5 million daily users and as you can imagine that's a lot of traffic, that's a lot of searching. The phrase index is the way we're able to make sure that PubMed performs for all of our 3.5 million users and provides you with the search results you're looking for. What this means is that when you're looking for an exact phrase, it needs to be included in PubMed's phrase index in order to search for it.

So an example that I got at the help desk recently was for this term ACU 193, which is-- OK well I'm not going to try and pronounce it, but this was something that someone emailed me about and it wasn't in the phrase index. And the way that you can know it's not in the phrase index is by checking, by typing this in here into the Advanced Search Builder and then click Show Index. And I've typed in my exact term and we can see that nothing is showing up here for 193. Maybe if I go back and just look for ACU. Now we can see that there are other similar terms with other numbers associated with it here, but still it's not in the phrase index. If I hadn't checked though, and I wanted to look for this term and I put that into PubMed with the double quotes and click Search, you'll get this message that says quoted phrase not found. Because this phrase isn't in PubMed's phrase index and the reason it's not in PubMed's phrase index is because it doesn't appear on any PubMed records yet.

And before I get to the specific question about MTIX, I'll just give a quick plug for our proximity searching feature. This is another way you can check for the terms that maybe aren't in the phrase index yet. You can use our proximity search with an N value of 0, and there's instructions for how to do this in the user guide if you want to try this yourself. So I'm looking in the title and abstract field for these terms with a distance of 0 in between each other and we know it doesn't appear in PubMed yet because we didn't retrieve any results with our proximity search here.

So when you get to this point, one thing I'll suggest is browsing the index like we just looked at on the advanced search page. Because the solution for this user is actually that generally this term doesn't have a space between it when it's represented in the literature and that's why they were encountering that issue. But if you run into that wall and you're not sure what to do next, feel free to write to the help desk. We always like to help.

As far as improvements to the phrase index using the MTIX algorithm, it's not something we're looking at, and the reason for that is because we already have an algorithm in place that evaluates new phrases in PubMed. We're constantly checking to see how frequently a phrase appears. We have some criteria that we use to look for that phrase and how often it should appear before it shows up in the phrase index. So we already have that algorithm in place, so we're not looking for anything to replace that. And the final plug that I will give here is that if you see a phrase, you find it using proximity and say like, hey, this appears on a citation already. Can you add this to the phrase index? We will manually add it as long as it appears on at least one citation. So again, write to the help desk if you're struggling with the phrase index or if you have a request for a phrase and we would be happy to help you.

MIKE: Before we let you go, somebody Jill, is asking **how do I contact the PubMed help desk?** And since you're already sharing your screen, I thought this would be a fantastic question.

AMANDA: Yes. Goodness, I don't think I've ever done it from this page. There we go. Thank you. Sorry. Use the Help button. It's at the bottom of all of our pages and that will take you to the NLM support center. And click this button here, it says Write to the help desk and you can provide a subject here. It definitely helps us if you tell us what NLM product you're writing about. So PubMed is a product, you can put that in the subject and that helps get your question routed to the right team. When you write about PubMed, your questions get sent to someone like me or any of my colleagues who are trained to answer PubMed questions specifically. And we love hearing from you whether it's a question or feedback. So please write. We'd love to hear from you.

MIKE: Excellent. Yeah, and I'll also point out that the NLM Support Center link which is on the bottom of all non-PubMed NLM websites also will get you to the same place. So if you're looking for help, either the help link at the bottom of the PubMed page or the NLM Support Center link at the bottom of any of our other web pages will get you here. You can write to the help desk. And again to echo what Amanda said, we really do appreciate hearing from you. These things just don't go into the ether. Many of the people who answer your questions are on this call right now and definitely want to hear more from you.

All right, we are running out of time. So I'm going to try to get to as many as I can in the last few minutes here. We had a couple of questions that were submitted ahead of time. So I want to make sure I get to answer some of those. Deborah, one of those questions was **how are publication types assigned to articles?**

DEBORAH: Hi, yes, publication types are assigned one of two ways. The publisher can submit them directly when they send in their citations and then our automated indexing algorithm also assigns citations. And both of those can have errors. Sometimes the publisher misuses or misunderstands a publication type and if it's that's a systematic error, we'll contact the publisher and try to get them to fix that. And then with automated indexing, that's one of the areas where we focus on curation. It's specifically on publication types, especially the high impact ones like clinical trials and systematic reviews. Curators look at every citation and we'll fix those if they need fixing.

MIKE: Excellent. Thank you for that. All right, just a few more minutes here. Get through a couple more questions if we can. Where was that question that I was going to look at? Oh, this is a great question. Alex, you can definitely answer this one. Tracy asks **what mechanisms are in place to deal with indexing issues that users encounter?**

ALEX: If you identify an indexing issue that you would like to tell us about, you can submit a report via form and the support desk. And I assume that we can put that link in the chat.

MIKE: Yes, somebody will get that link in the chat.

ALEX: We do read and respond to every report that we get via customer service. And many of those reports have been very useful to us in identifying things that have been systematic errors and allowing us to improve the performance of MTIA or to figure out better ways to kind of address user concerns. So we welcome that feedback very much.

MIKE: Yes, thank you so much for that. And I know that we have a number of you who have questions that are outstanding, that have not yet been answered. We're going to do our best to get whatever questions we can answered and we'll try to get some answers posted for the rest of you folks after the session. Here we go. For Amanda, Iris asks, **will Best Match work more effectively with this new indexing algorithm?**

AMANDA: You know, that's a really great question and I don't think you have a definitive answer to it, but my assumption is the better the indexing, the better Best Match can perform because Best Match is based off of the text on the records and how well that matches to your search terms. I think I've mentioned in the past, there are some papers published about how Best Match works. But if you have specific questions about it, I'll plug the help desk again, write and we can talk to a developer and maybe hopefully get you a more definitive answer.

MIKE: Thank you for that. I'm just trying to see if I can get one more quick one in here. Alex, since you're already on camera, Sarah's asking **does MTIX assign qualifiers and if so do we know how well it performs?**

ALEX: MTIX does assign qualifiers, and the evaluation for qualifiers is difficult because it's hard to say when a qualifier is assigned to a term that wasn't in the original indexing. So when we're making these comparisons and we compare like the descriptors of the SCRs, we can sort of compare them 1:1. And then we can compare the subheadings as qualifiers on terms that were in the original indexing. But then for any terms that MTIX has added extra, if it's also added qualifiers on to those terms, then we have to count them all as wrong even though it turns out when we look at those sort of individually, they might not be wrong. So it gets very confusing to try and present the metrics for the subheadings, which is one of the reasons I didn't cover them.

I can say that the performance there, it's certainly not as good. And this will actually address, I think another question that we had coming in, which is that the reliability of how consistently the indexers have used a certain term has a huge impact on how well the algorithm performs with those terms. So one of the reasons that we have such high performance with checktags is because checktags are used extremely consistently by the indexers and the same is true of publication types in general as well. And so then when we look at the subheadings, that is the area where the indexers have the least amount of consistency from indexer to indexer, the subheadings that they would choose to put on a particular term. And so MTIX also has fairly high inconsistency on checktags when we look at it that way-- or sorry, not checktags, on subheadings when we look at it that way. So, the performance on subheadings isn't great, but it's difficult to give exact numbers on it because the numbers end up being very complicated. So we didn't want to present them.

MIKE: Thanks. Yeah, a lot of this stuff does come down to being a little bit more complicated to evaluate them then it might seem like at first. So thank you for that and for the work that you've done. It is past the time that we are supposed to have wrapped up. So I'm going to wrap us up now and we're going to make sure that we save off the rest of the Q&A and get some answers posted for questions that were not addressed during the session.

Again, I want to thank all of our panelists. Thanks Alex for a great presentation and thanks to all of you who are attending for your wonderful questions. It was very lively. There's clearly a lot of interest here.

We are continuing to have these Office Hours sessions. We will have more. Usually we have them every other month or so and stay tuned to be NLM Technical Bulletin to see more about those.