MEDLINE Indexing Update

The MTIX Algorithm

Alex Sticco

Thursday, March 21, 2024

Overview

Automation in Indexing

- What is it?
- What is changing?

New Algorithm Technology

- How does it work?
- How is it different?

Evaluation and Performance

- How do we evaluate performance?
- How well does it work?

Overview 1

Automation in Indexing

- What is it?
- What is changing?

Automated MEDLINE Indexing

- Automated all MEDLINE indexing in 2022
  - Indexed a backlog of ~850,000 citations
  - Indexing lag was 1-3 months, now just 1 day!

- Updating automation in 2024
  - New, improved indexing algorithm

Terminology

MTIA =

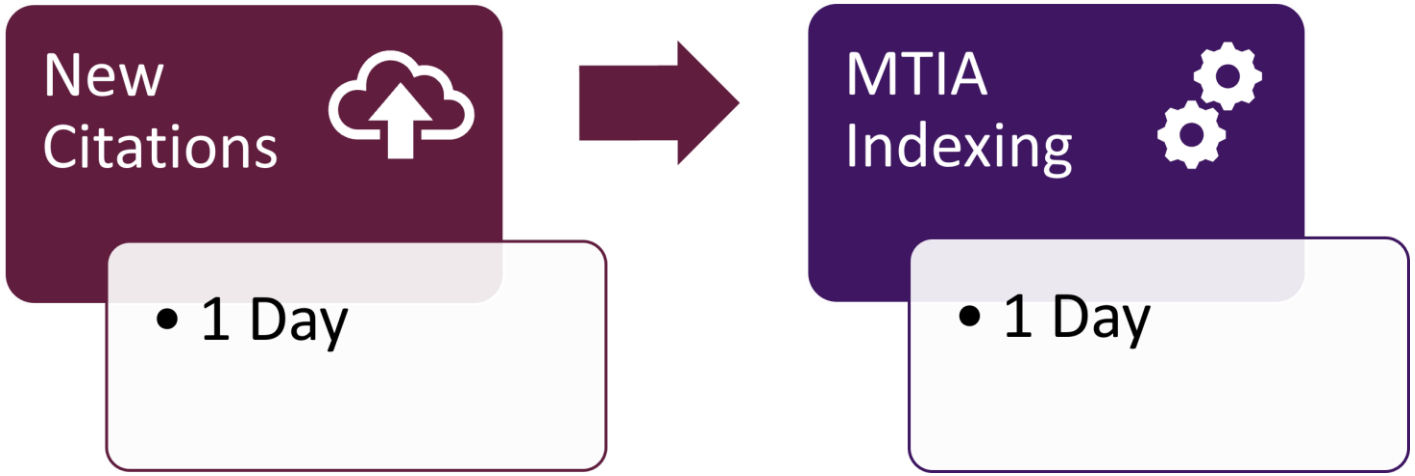
Current Algorithm

MTIX =

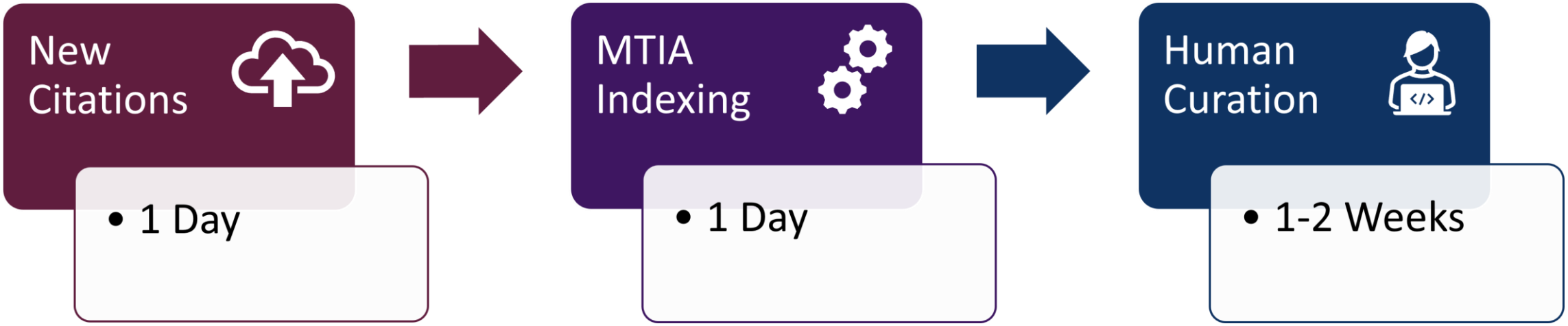
New Algorithm

MEDLINE Indexing Pipeline

MEDLINE Indexing Pipeline



MEDLINE Indexing Pipeline



MEDLINE Indexing Pipeline

Not Changing

Not Changing

- Indexing Scope

Not Changing

- Indexing Scope
  - MEDLINE only
  - No preprints

Not Changing

- Indexing Scope
  - MEDLINE only
  - No preprints

- Indexing from the Title and Abstract

Not Changing

- Indexing Scope
  - MEDLINE only
  - No preprints

- Indexing from the Title and Abstract
  - Full text license issues

Not Changing

- Indexing Scope
  - MEDLINE only
  - No preprints

- Indexing from the Title and Abstract
  - Full text license issues
  - Model input limits

Overview

Automation in Indexing

- What is it?
- What is changing?

Overview

Automation in Indexing

- What is it?
- What is changing?

New Algorithm Technology

- How does it work?
- How is it different?

MTIA vs. MTIX: Technology

MTIA	MTIX

MTIA vs. MTIX: Technology

MTIA	MTIX
• Dictionary and Rules	

MTIA vs. MTIX: Technology

MTIA	MTIX
• Dictionary and Rules • Humans create rules	

MTIA vs. MTIX: Technology

MTIA

- Dictionary and Rules
- Humans create rules

MTIX

- Machine Learning (AI)

MTIA vs. MTIX: Technology

MTIA

- Dictionary and Rules
- Humans create rules

MTIX

- Machine Learning (AI)
- Creates its own rules

MTIA vs. MTIX: Technology

MTIA

- Dictionary and Rules
- Humans create rules
- All heuristics

MTIX

- Machine Learning (AI)
- Creates its own rules
- All statistics

MTIA vs. MTIX: Performance

MTIA

MTIX

MTIA vs. MTIX: Performance

MTIA

MTIX

MTIA vs. MTIX: Performance

MTIA

- Limited to literal matches

MTIX

- Understands indirect and complex constructions

MTIA vs. MTIX: Performance

MTIA

- Limited to literal matches
- Makes contextual errors

MTIX

- Understands indirect and complex constructions
- Understands context

MTIA vs. MTIX: New MeSH Terms

MTIA

MTIX

MTIA vs. MTIX: New MeSH Terms

MTIA

MTIX

- New terms just like old terms

MTIA vs. MTIX: New MeSH Terms

MTIA

MTIX

- New terms just like old terms
- Performance depends on:
  - Comprehensive synonyms
  - Exact words in text

MTIA vs. MTIX: New MeSH Terms

MTIA

- New terms just like old terms
- Performance depends on:
  - Comprehensive synonyms
  - Exact words in text

MTIX

- Needs new training data

MTIA vs. MTIX: New MeSH Terms

MTIA

- New terms just like old terms
- Performance depends on:
  - Comprehensive synonyms
  - Exact words in text

MTIX

- Needs new training data
- Temporarily “MTIA-lite” system

Overview

Automation in Indexing

- What is it?
- What is changing?

New Algorithm Technology

- How does it work?
- How is it different?

Overview

Automation in Indexing

- What is it?
- What is changing?

New Algorithm Technology

- How does it work?
- How is it different?

Evaluation and Performance

- How do we evaluate performance?
- How well does it work?

Evaluation Methods and Metrics

Method:

Comparison to Manual Indexing

Metrics:

If you are interested in…	The metric you need is…
How many terms did MTIA/MTIX miss?	Recall
How many extra incorrect terms did MTIA/MTIX add?	Precision
A single number that combines precision and recall	F1

Comparison to Manual Indexing

Strengths:

- Relevant
- Simple and fast
- Abundant, diverse data for testing at scale

Comparison to Manual Indexing

Weaknesses:

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn’t technically a “gold standard”

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary
  - Variable secondary concepts

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary
  - Variable secondary concepts
  - Human error

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary
  - Variable secondary concepts
  - Human error
- Errors are not all equally important

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary
  - Variable secondary concepts
  - Human error
- Errors are not all equally important
  - Impact on searchers

Comparison to Manual Indexing

Weaknesses:

- Manual indexing isn't technically a "gold standard"
  - Variable vocabulary
  - Variable secondary concepts
  - Human error
- Errors are not all equally important
  - Impact on searchers
  - Degree of "wrongness

Upshot

- Many MTIX “errors” are actually:

Upshot

- Many MTIX “errors” are actually:
  - Improvements or corrections

Upshot

- Many MTIX “errors” are actually:
  - Improvements or corrections
  - Alternative ways to represent the same concepts

Upshot

- Many MTIX “errors” are actually:
  - Improvements or corrections
  - Alternative ways to represent the same concepts
  - Relevant but unimportant concepts

Upshot

- Many MTIX “errors” are actually:
  - Improvements or corrections
  - Alternative ways to represent the same concepts
  - Relevant but unimportant concepts
  - Real errors, but with low impact

Upshot

- Many MTIX “errors” are actually:
  - Improvements or corrections
  - Alternative ways to represent the same concepts
  - Relevant but unimportant concepts
  - Real errors, but with low impact

- But…
  - We can’t determine *which* errors are which, at scale

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

Missed Terms

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

- Truly Incorrect

Missed Terms

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

- Truly Incorrect
- Improvement/Equal

Missed Terms

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms	Missed Terms

- Truly Incorrect
- Improvement/Equal
- OK: not as good
- OK: unimportant

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

- Truly Incorrect
- Improvement/Equal
- OK: not as good
- OK: unimportant

Missed Terms

- Truly Missed

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

- Truly Incorrect
- Improvement/Equal
- OK: not as good
- OK: unimportant

Missed Terms

- Truly Missed
- Improved/Equal

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

Extra Terms

- Truly Incorrect
- Improvement/Equal
- OK: not as good
- OK: unimportant

Missed Terms

- Truly Missed
- Improved/Equal
- Partially Covered

MTIX Error Evaluation

- 185 manually indexed articles from 2022
- Indexers evaluated errors

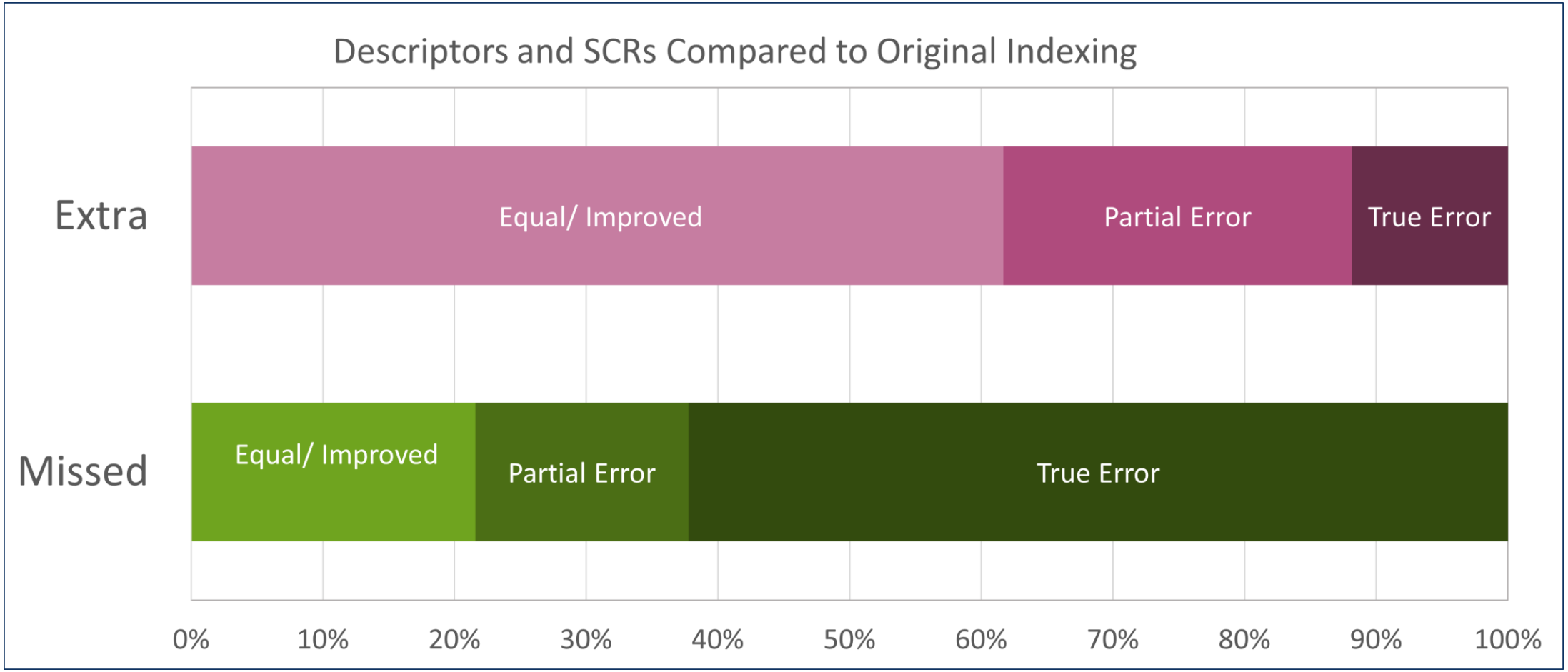
Extra Terms

- Truly Incorrect
- Improvement/Equal
- OK: not as good
- OK: unimportant

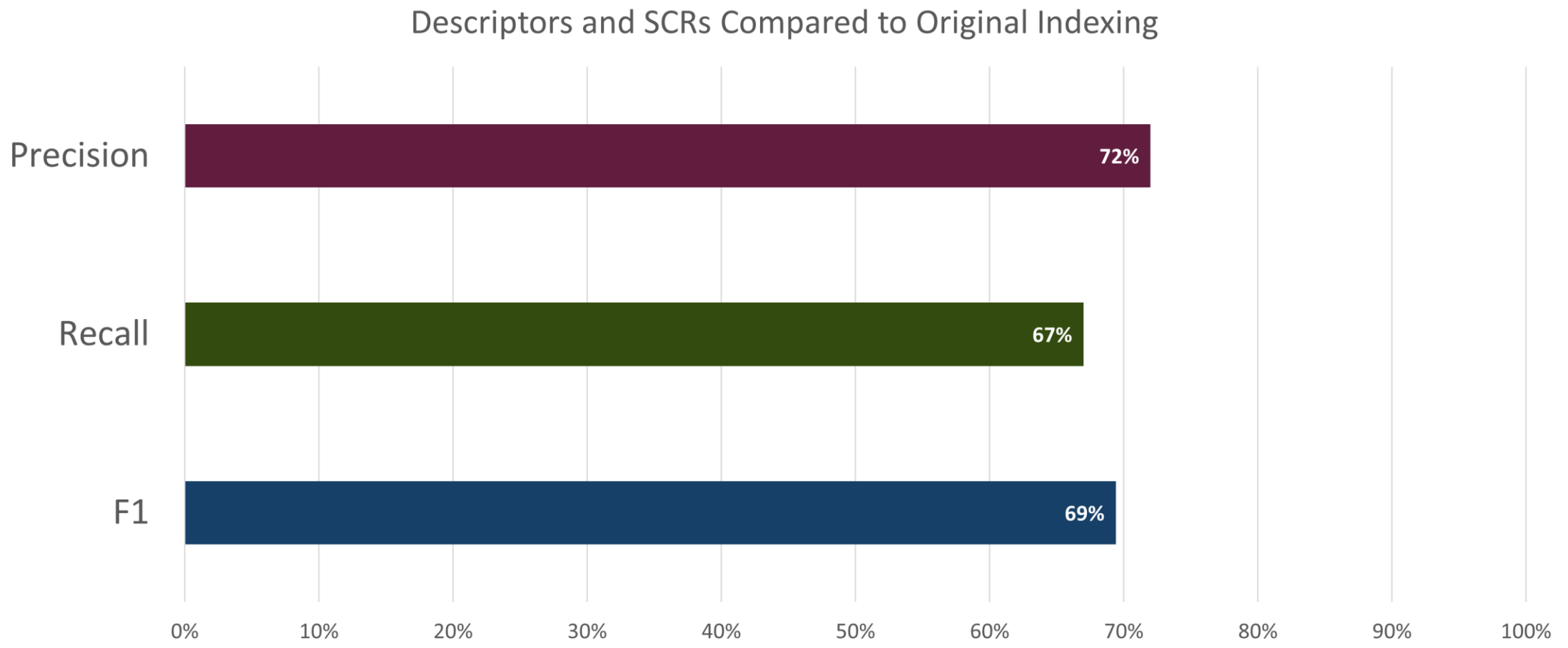
Missed Terms

- Truly Missed
- Improved/Equal
- Partially Covered
- Originally wrong

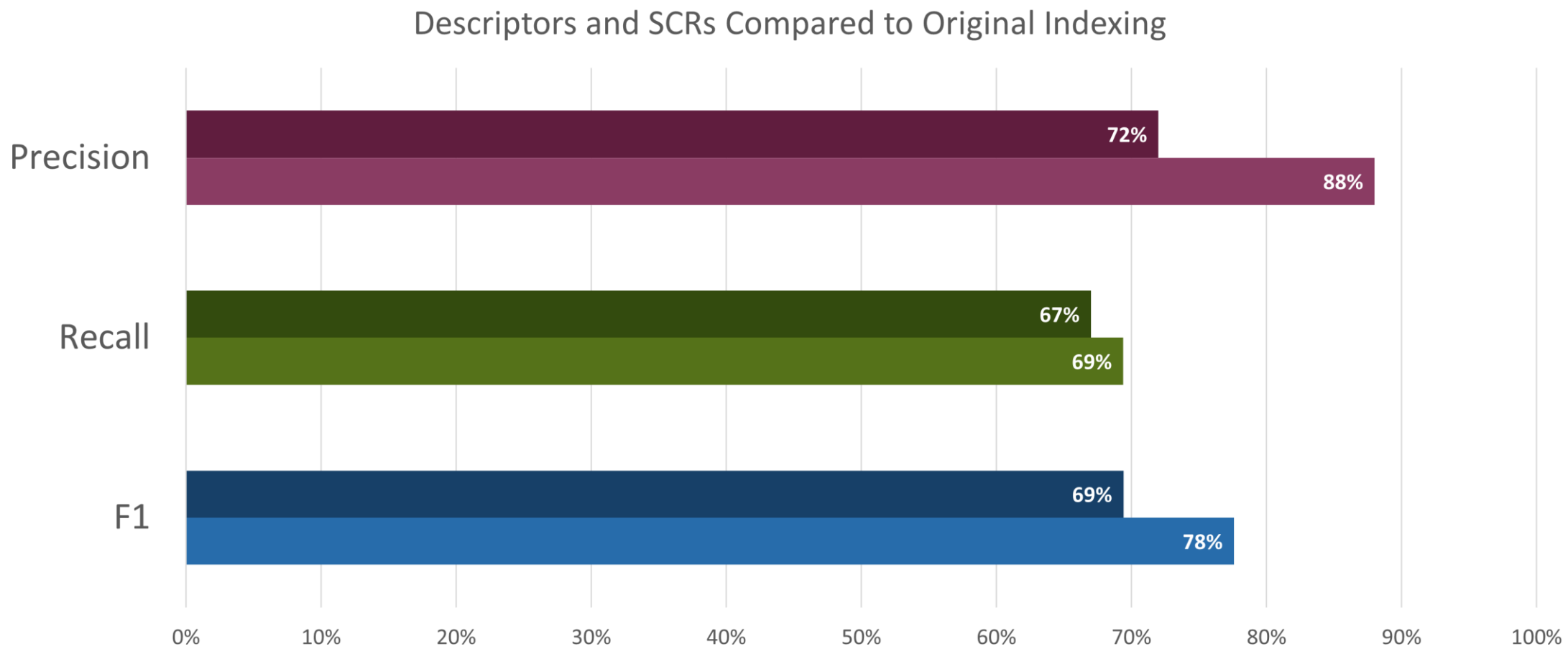
MTIX Error Evaluation Results



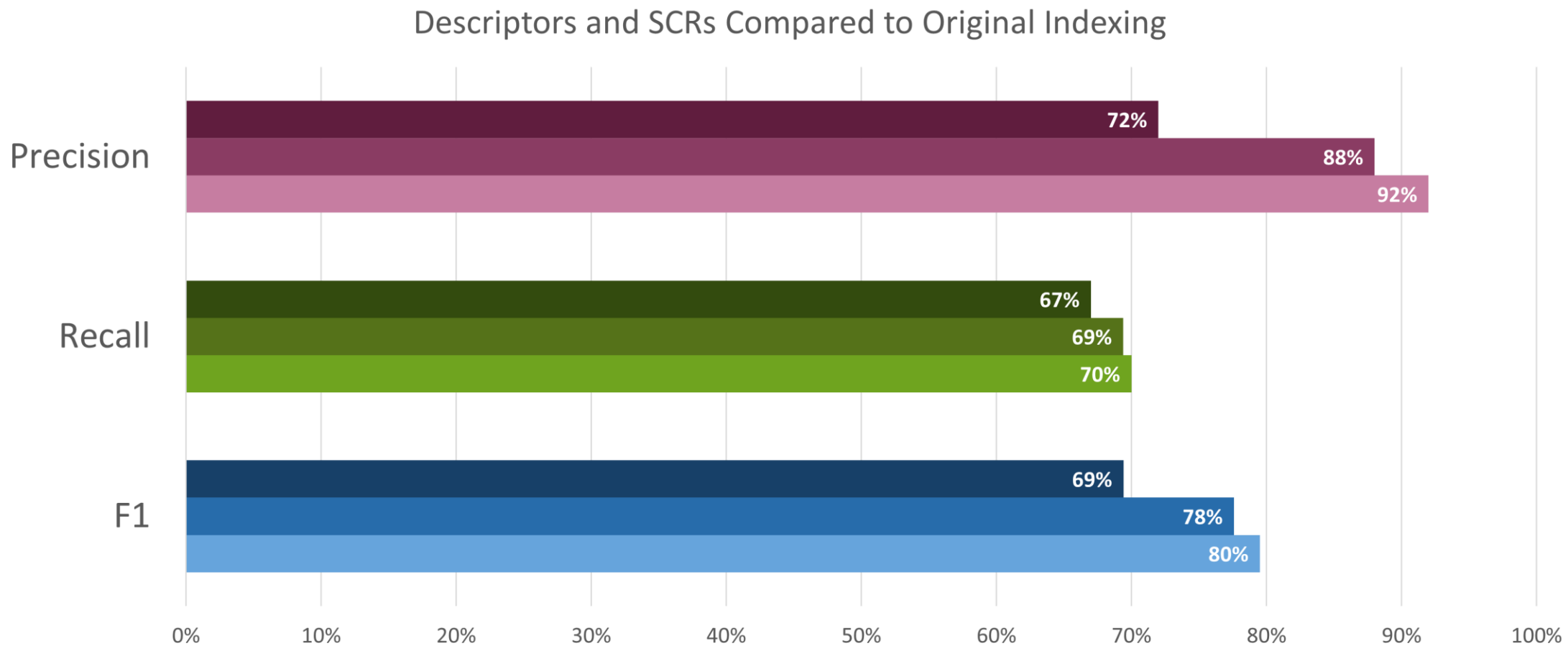
MTIX Raw Metrics



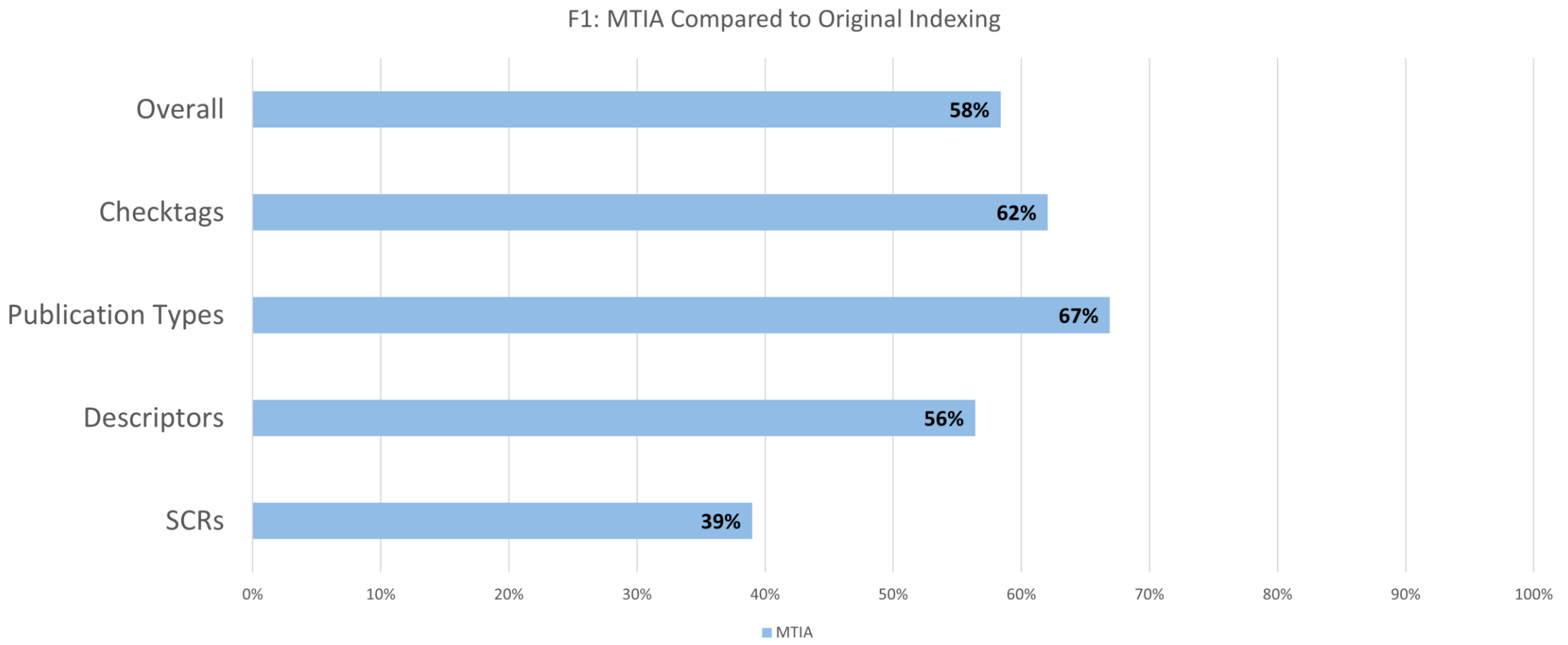
MTIX Raw vs Adjusted Metrics



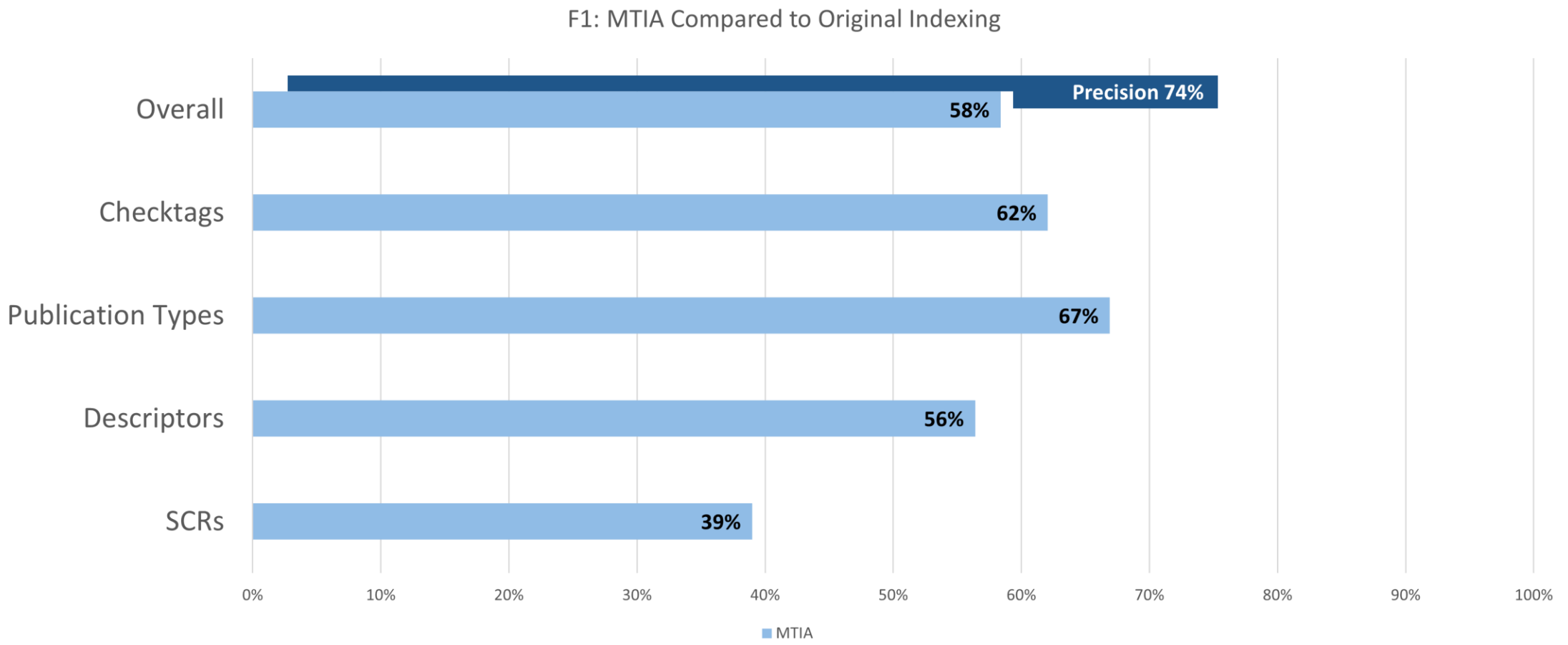
MTIX Raw vs Adjusted Metrics



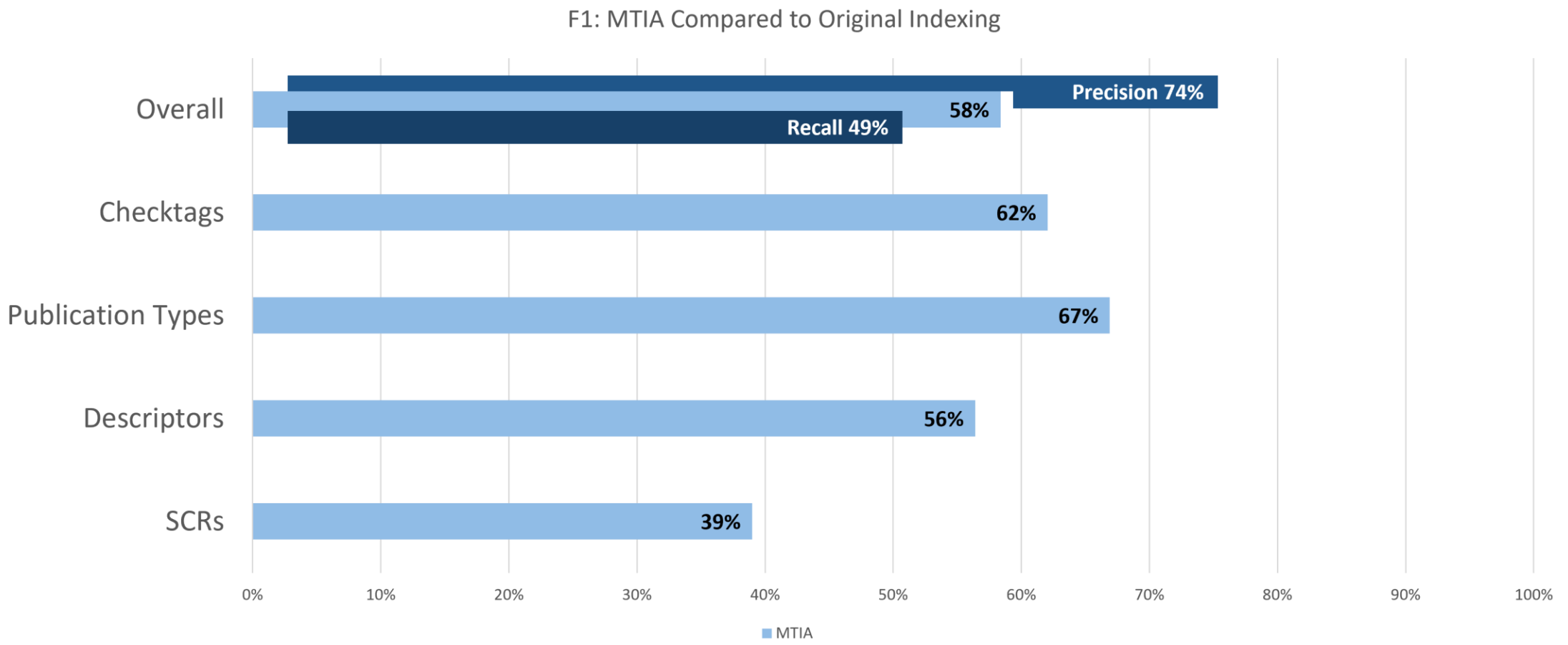
MTIA F1 Scores



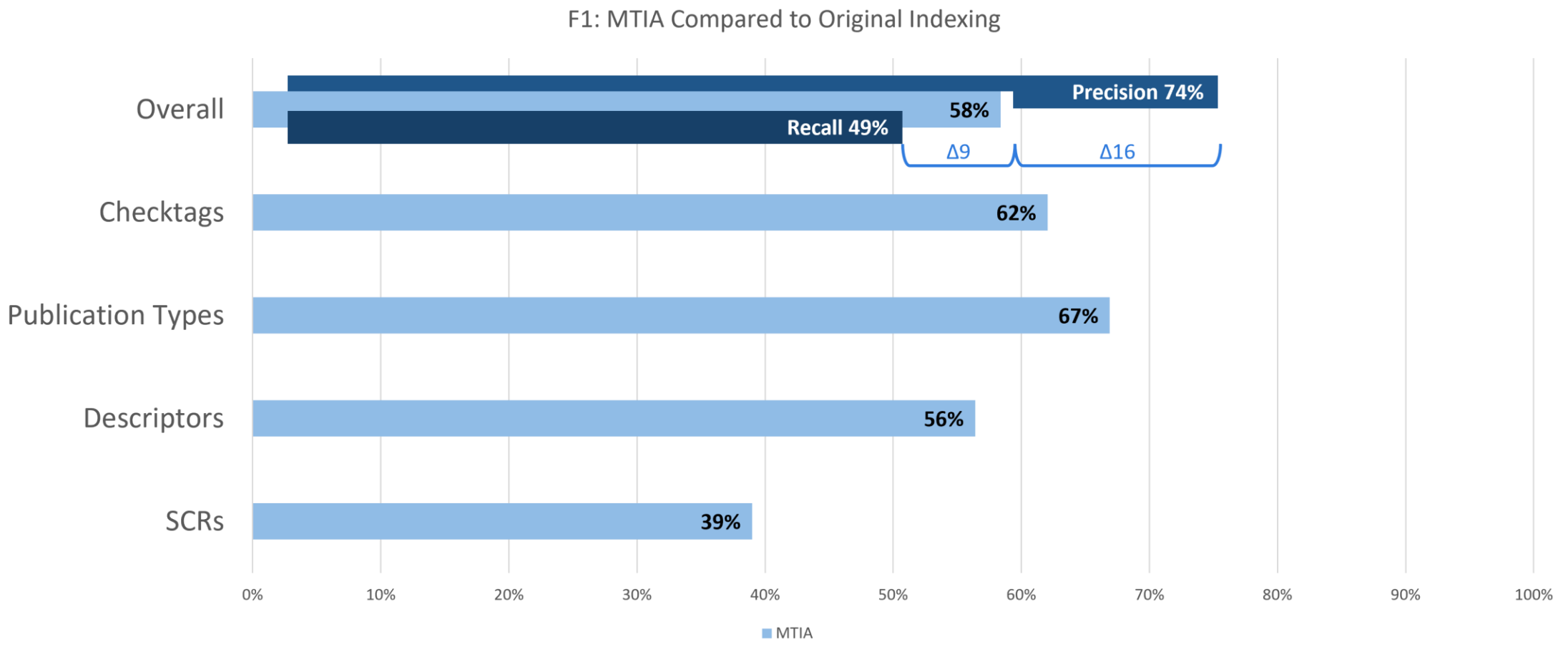
MTIA F1 Scores



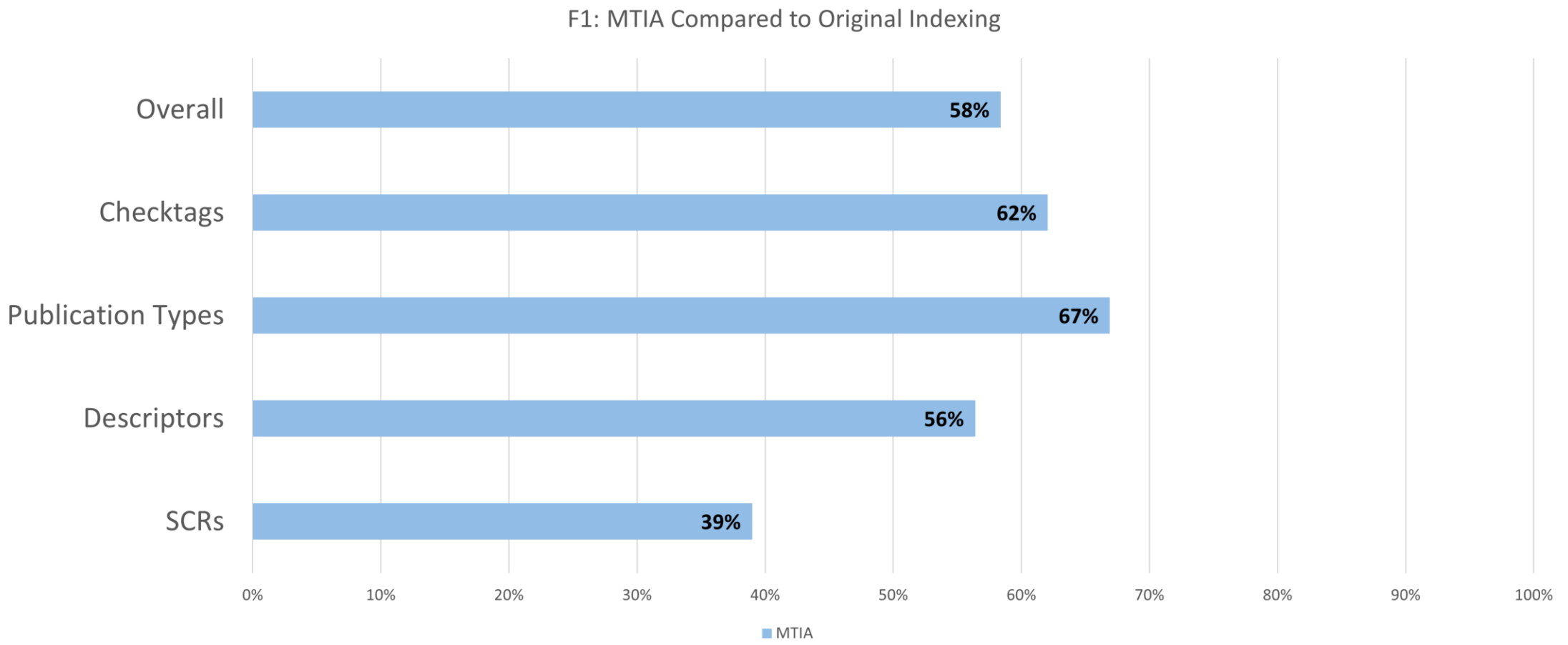
MTIA F1 Scores 2



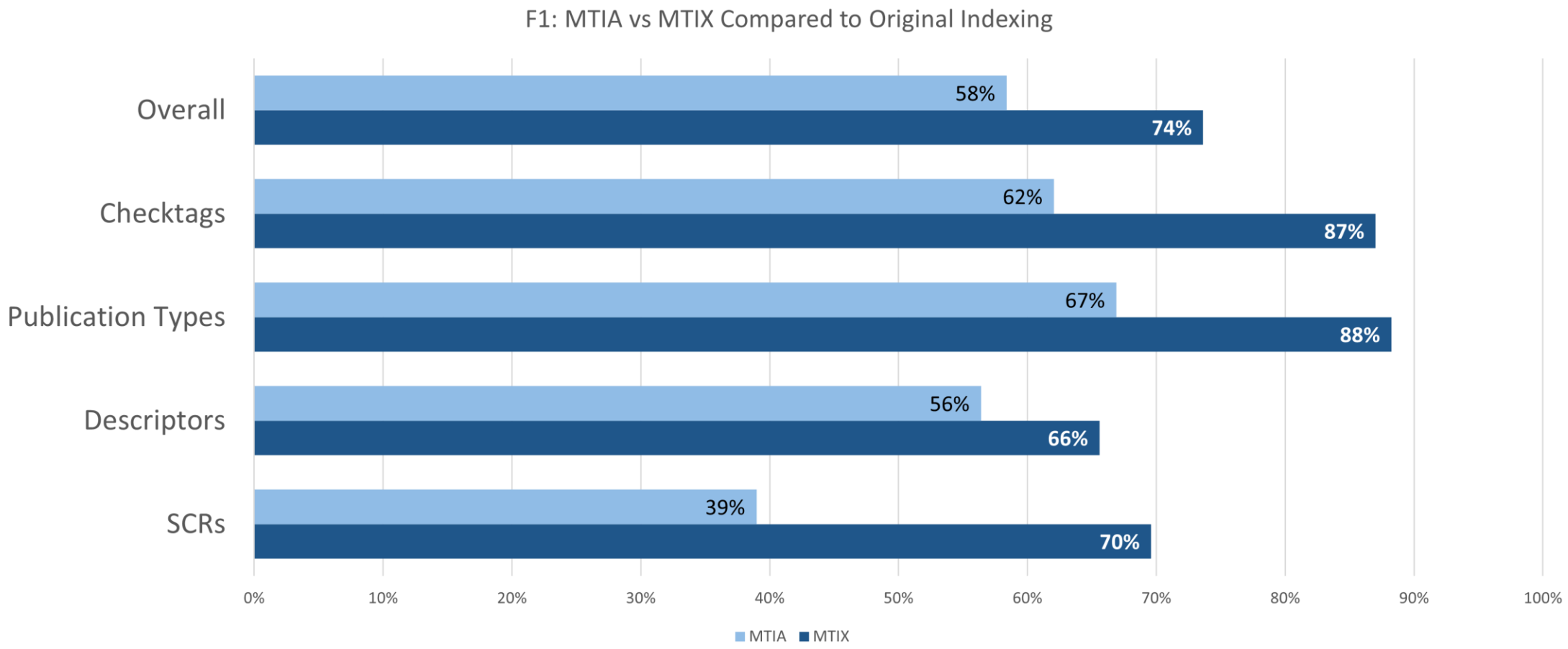
MTIA F1 Scores



MTIA F1 Scores



MTIA vs MTIX F1 Scores



MTIX Checktag Performance

	Humans	Animals	Female	Male	Pregnancy	Infant	Infant, Newborn	Child	Child, Preschool	Adolescent	Young Adult	Adult	Middle Aged	Aged	Aged, 80 and Over
Recall	98%	95%	92%	93%	93%	72%	61%	82%	78%	62%	64%	87%	91%	86%	58%
Precision	95%	90%	85%	84%	82%	76%	71%	81%	78%	74%	64%	75%	81%	79%	68%
F1	**96%**	**92%**	**89%**	**88%**	**87%**	**74%**	**66%**	**81%**	**78%**	**68%**	**64%**	**81%**	**86%**	**83%**	**63%**

MTIX Checktag Performance

	Humans	Animals	Female	Male	Pregnancy	Infant	Infant, Newborn	Child	Child, Preschool	Adolescent	Young Adult	Adult	Middle Aged	Aged	Aged, 80 and Over
Recall	98%	95%	92%	93%	93%	72%	61%	82%	78%	62%	64%	87%	91%	86%	58%
Precision	95%	90%	85%	84%	82%	76%	71%	81%	78%	74%	64%	75%	81%	79%	68%
F1	**96%**	**92%**	**89%**	**88%**	**87%**	**74%**	**66%**	**81%**	**78%**	**68%**	**64%**	**81%**	**86%**	**83%**	**63%**

MTIX Checktag Performance

	Humans	Animals	Female	Male	Pregnancy	Infant	Infant, Newborn	Child	Child, Preschool	Adolescent	Young Adult	Adult	Middle Aged	Aged	Aged, 80 and Over
Recall	98%	95%	92%	93%	93%	72%	61%	82%	78%	62%	64%	87%	91%	86%	58%
Precision	95%	90%	85%	84%	82%	76%	71%	81%	78%	74%	64%	75%	81%	79%	68%
F1	**96%**	**92%**	**89%**	**88%**	**87%**	**74%**	**66%**	**81%**	**78%**	**68%**	**64%**	**81%**	**86%**	**83%**	**63%**

MTIX Checktag Performance

	Humans	Animals	Female	Male	Pregnancy	Infant	Infant, Newborn	Child	Child, Preschool	Adolescent	Young Adult	Adult	Middle Aged	Aged	Aged, 80 and Over
Recall	98%	95%	92%	93%	93%	72%	61%	82%	78%	62%	64%	87%	91%	86%	58%
Precision	95%	90%	85%	84%	82%	76%	71%	81%	78%	74%	64%	75%	81%	79%	68%
F1	**96%**	**92%**	**89%**	**88%**	**87%**	**74%**	**66%**	**81%**	**78%**	**68%**	**64%**	**81%**	**86%**	**83%**	**63%**

MTIX Checktag Performance

	Humans	Animals	Female	Male	Pregnancy	Infant	Infant, Newborn	Child	Child, Preschool	Adolescent	Young Adult	Adult	Middle Aged	Aged	Aged, 80 and Over
Recall	98%	95%	92%	93%	93%	72%	61%	82%	78%	62%	64%	87%	91%	86%	58%
Precision	95%	90%	85%	84%	82%	76%	71%	81%	78%	74%	64%	75%	81%	79%	68%
F1	**96%**	**92%**	**89%**	**88%**	**87%**	**74%**	**66%**	**81%**	**78%**	**68%**	**64%**	**81%**	**86%**	**83%**	**63%**

Up Next…

Up Next…

- Generate training data for new MeSH terms

Up Next…

- Generate training data for new MeSH terms
- Use Pubmed search data to guide development

Up Next…

- Generate training data for new MeSH terms
- Use Pubmed search data to guide development
- Keep an eye on developments in AI, especially SLM and LLM

We Want to Hear From You!

Live Participants

Submit questions to the Q&A chat

Watching Later?

(888) FIND-NLM

support.nlm.nih.gov

nlm.nih.gov/socialmedia

Thanks for Attending!