The NLM Indexing Initiative: Current Status and Role in Improving Access to Biomedical Information

A Report to the Board of Scientific Counselors

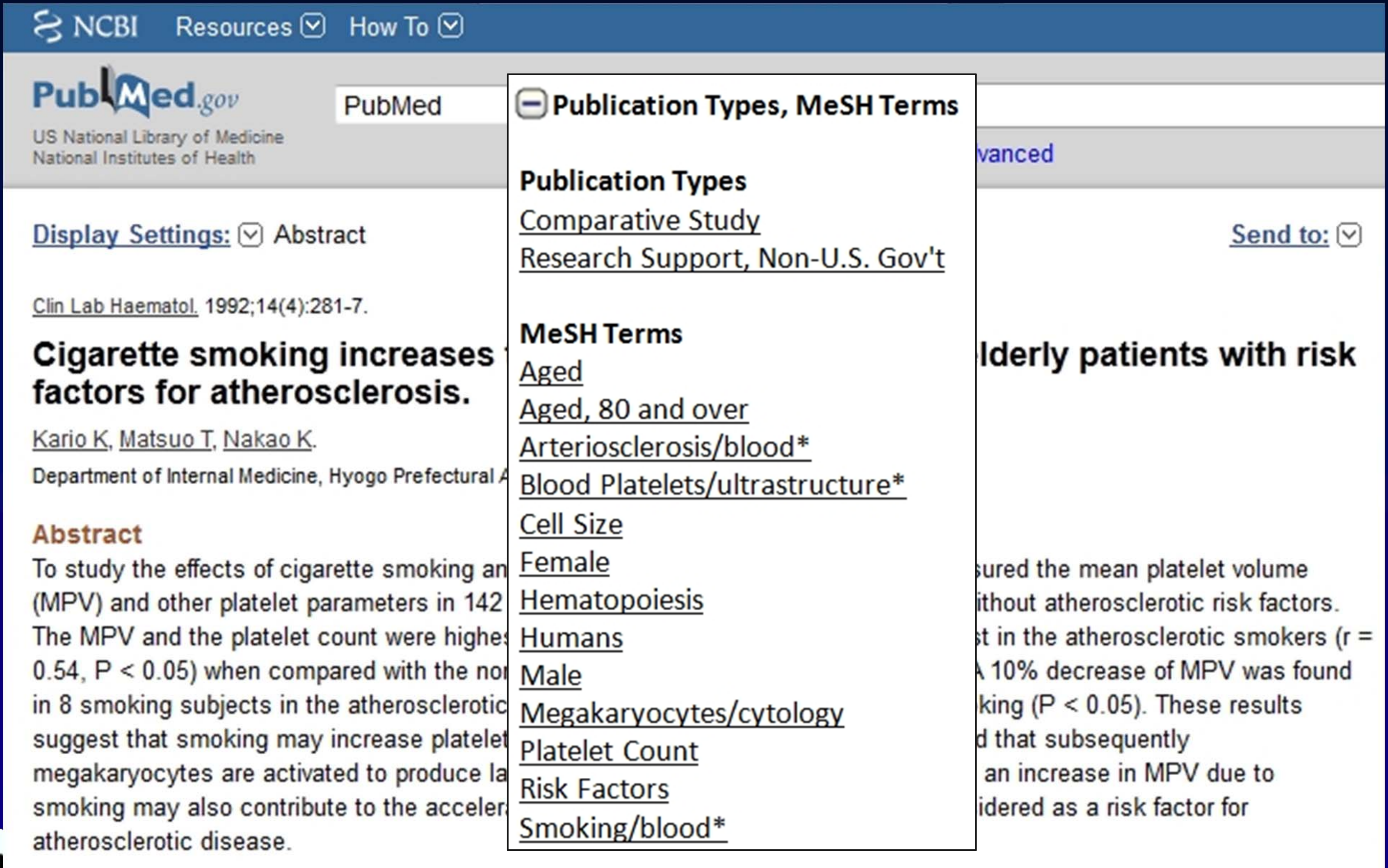
April 5, 2012

Alan R. Aronson (Principal Investigator)
James G. Mork
François-Michel Lang
Willie J. Rogers
Antonio J. Jimeno-Yepes
J. Caitlin Sticco

Outline

- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

MEDLINE Citation Example



NCBI
Resources
How To
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
vanced
Display Settings: Abstract
Send to:

Clin Lab Haematol. 1992;14(4):281-7.

Cigarette smoking increases … **lderly patients with risk factors for atherosclerosis.**

Kario K, Matsuo T, Nakao K.

Department of Internal Medicine, Hyogo Prefectural A

Abstract

To study the effects of cigarette smoking an … sured the mean platelet volume (MPV) and other platelet parameters in 142 … ithout atherosclerotic risk factors. The MPV and the platelet count were highe … t in the atherosclerotic smokers (r = 0.54, P < 0.05) when compared with the no … A 10% decrease of MPV was found in 8 smoking subjects in the atherosclerotic … king (P < 0.05). These results suggest that smoking may increase platelet … d that subsequently megakaryocytes are activated to produce la … an increase in MPV due to smoking may also contribute to the acceler … idered as a risk factor for atherosclerotic disease.

Publication Types, MeSH Terms

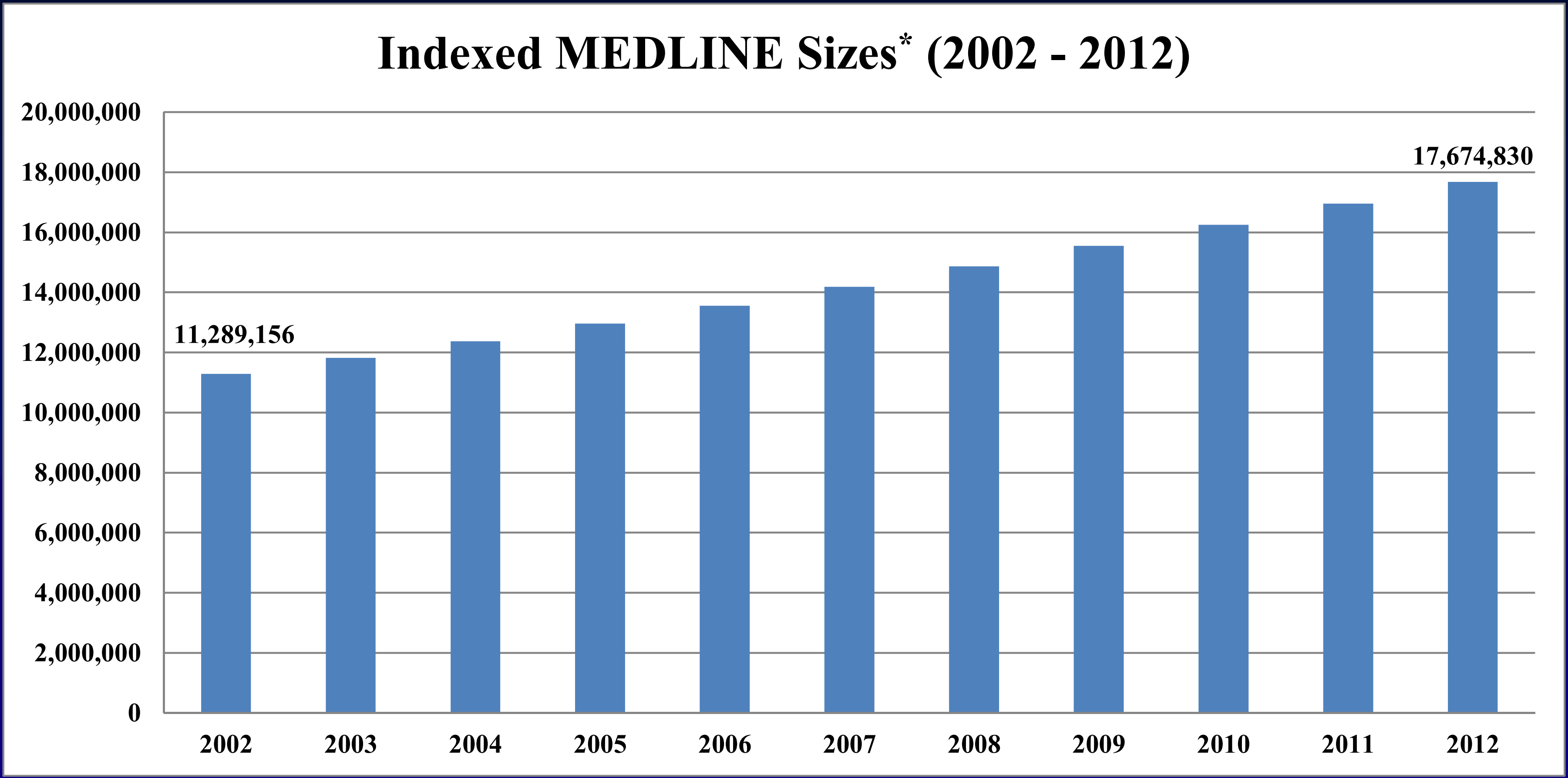
Publication Types

Comparative Study
Research Support, Non-U.S. Gov't

MeSH Terms

Aged
Aged, 80 and over
Arteriosclerosis/blood*
Blood Platelets/ultrastructure*
Cell Size
Female
Hematopoiesis
Humans
Male
Megakaryocytes/cytology
Platelet Count
Risk Factors
Smoking/blood*

Introduction - Growth in MEDLINE



* MEDLINE Baseline less OLDMEDLINE and PubMed-not-MEDLINE

The NLM Indexing Initiative (II)

- The need for MEDLINE indexing support
  - Increasing demand/costs for indexing in light of
  - Flat budgets
- One solution: creation of the NLM Indexing Initiative in 1996 resulting in NLM Medical Text Indexer (MTI)
- The Indexing Initiative today:
  - Identification of problems or needs followed by subsequent research
  - Production of MTI recommendations and other indexing
  - Opportunities for training and collaboration

Medical Informatics Training Program Fellows

- Antonio J. Jimeno-Yepes, Postdoctoral Fellow: 2010-
- J. Caitlin Sticco, Library Associate Fellow: 2011-
- Bridget T. McInnes, Postgraduate Fellow: 2008
  PhD in 2009
  Current affiliation: Securboration
- Aurélie Névéol, Postdoctoral Fellow: 2006-2008
  Current affiliation: NCBI
- Marc Weeber, Postgraduate Fellow: 2000
  PhD in 2001
  Current affiliation: Personalized Media

II Highlights from 2008

- Subheading attachment (Aurélie Névéol)
- Full text experiments (Cliff Gay)
- Initial Word Sense Disambiguation (WSD) method based on Journal Descriptor (JD) Indexing (Susanne Humphrey)
  - The *Journal of Cardiac Surgery* has JDs
  - 'Cardiology' and
  - 'General Surgery'

II Accomplishments since 2008

- The inauguration of MTI as a first-line indexer (MTIFL)
- Downloadable releases of MetaMap, most recently for Windows XP/7
- Significant improvement in MTI's performance due to
  - Technical improvements to MetaMap and MTI, but even more to
  - Close collaboration with LO Index Section
- More WSD methods with better performance
- The development of Gene Indexing Assistant (GIA)

Outline

- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

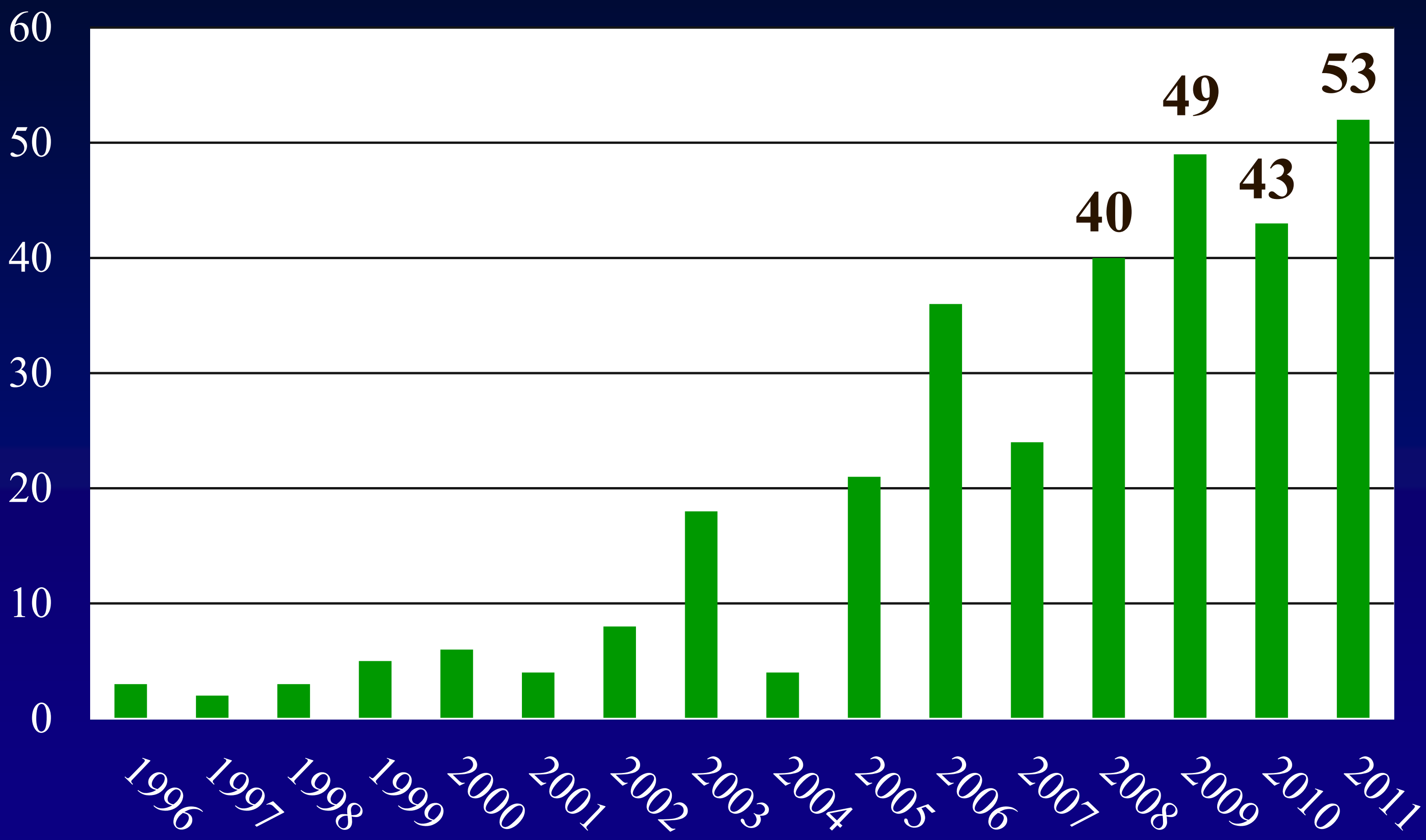
MetaMap - Overview

- Purpose
- Foundations
- Complexity
- Processing Example
- Challenge of UMLS Metathesaurus Growth
- Significant New Features

MetaMap - Purpose

- Named-entity recognition
- Identify UMLS Metathesaurus concepts in text
- Important and difficult problem
- MetaMap's dual role:
  - Local: Critical component of NLM's Medical Text Indexer (MTI)
  - Global: Pre-eminent biomedical concept-identification application

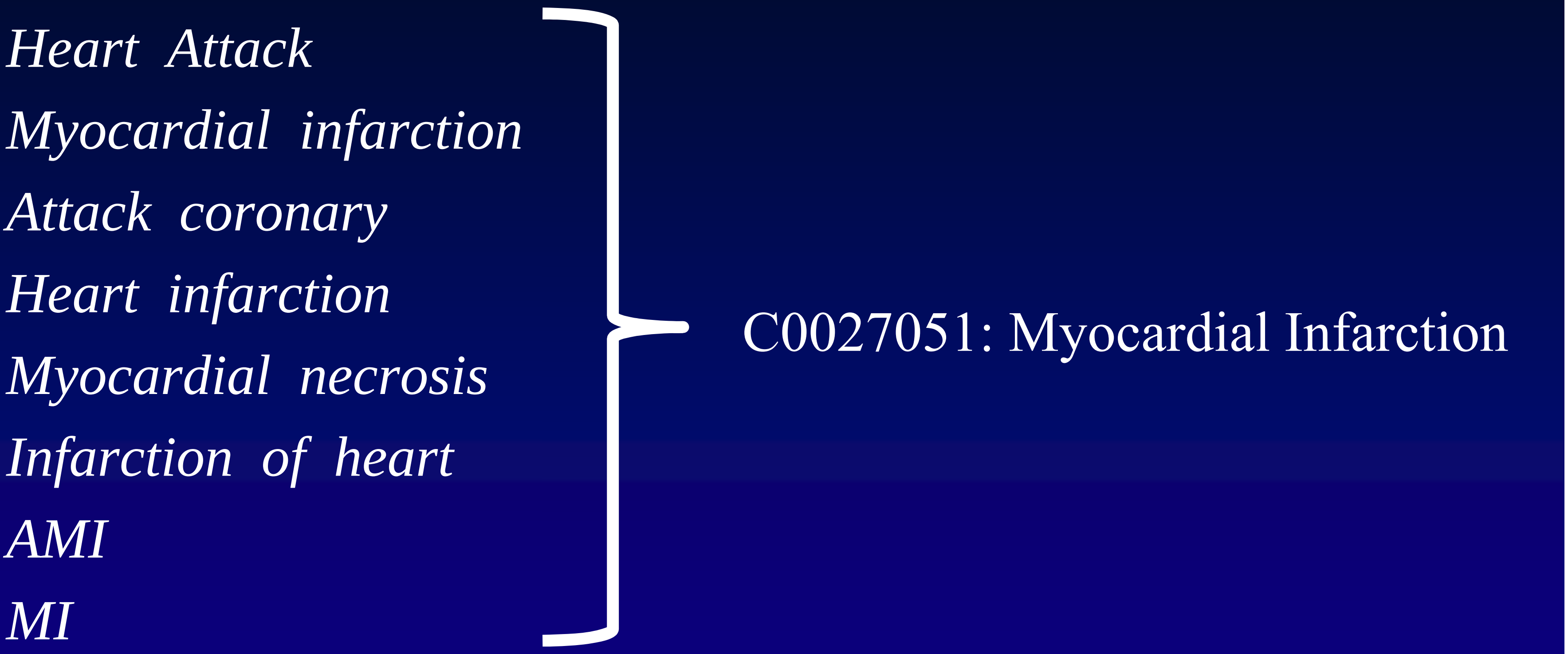
“MetaMap” in PubMed Central



MetaMap - Foundations

- Knowledge-intensive approach
- Natural Language Processing (NLP)
- Emphasize thoroughness over efficiency
- However…efficiency is still important!

Complexity of Language - Synonymy



Heart Attack

Myocardial infarction

Attack coronary

Heart infarction

Myocardial necrosis

Infarction of heart

AMI

MI

C0027051: Myocardial Infarction

Complexity of Language - Ambiguity

cold

C0009264: Cold Temperature

C0234192: Cold Sensation

C0009443: Common Cold

Ambiguity resolved by Word Sense Disambiguation

```
C0180860: Filters                   [mnob]
C0581406: Optical filter            [medd]
C1522664: filter information process [inpr]
C1704449: Filter (function)         [cnce]
C1704684: Filter Device Component   [medd]
```

Meta	Metathesau	Metathesauru	UMLS Semantic Type
909	C0080306:	Inferior Vena Cava Filter	[medd]
804	C0180860:	Filter	[mnob]
804	C0581406:	Filter	[medd]
804	C1522664:	Filter	[inpr]
804	C1704449:	Filter	[cnce]
804	C1704684:	Filter	[medd]
			[medd]
			[blor]
673	C0042460:	Vena caval	[bpoc]
637	C0038257:	Stent	[medd]
637	C1705817:	Stent	[medd]
637	C0447122:	Vena	[bpoc]

```
C0038257: Stent, device           [medd]
C1705817: Stent Device Component  [medd]
```

MetaMap - Processing Example

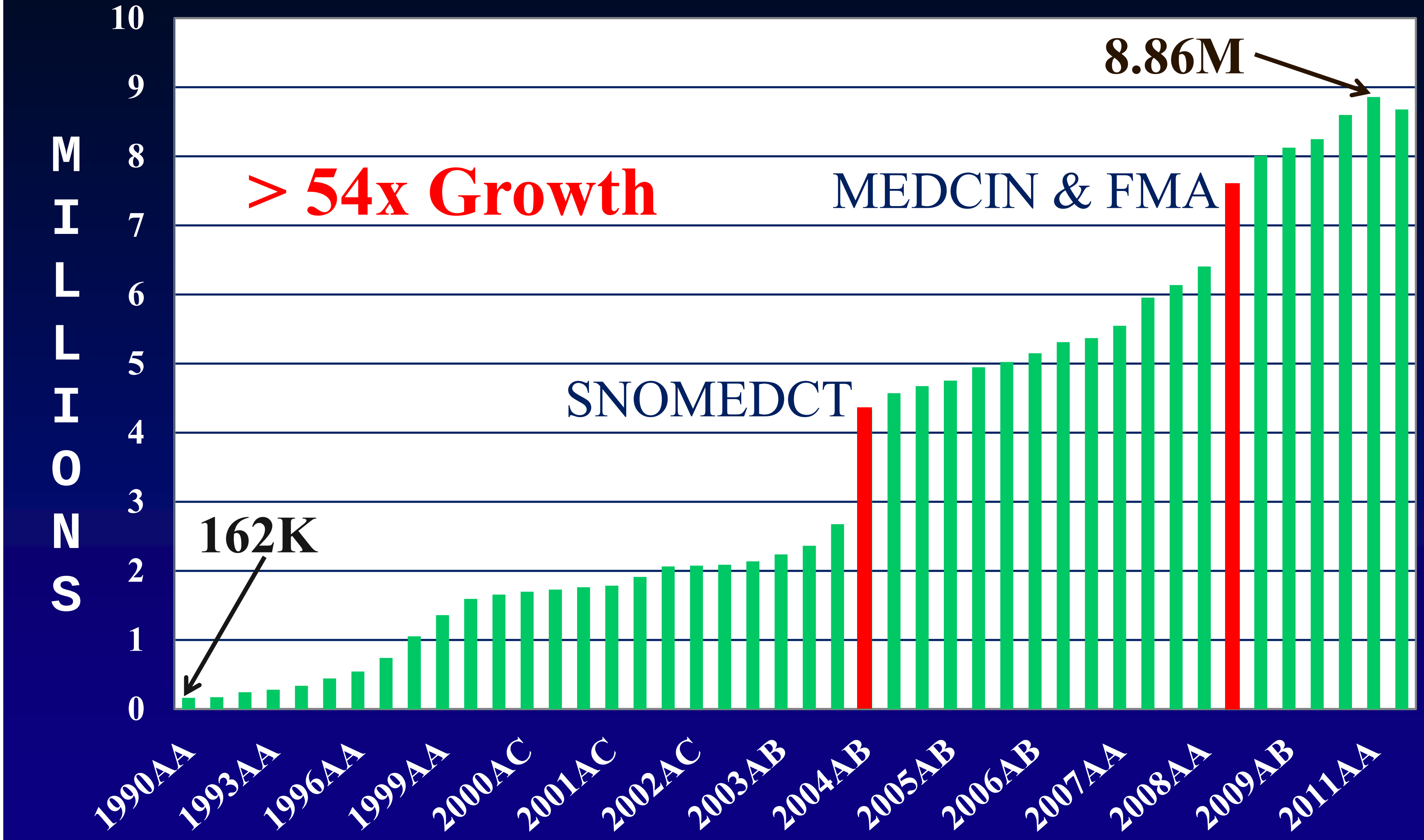
Inferior vena caval stent filter

Final Mappings (subsets of candidate sets):

```
Meta Mapping (911)
909  C0080306: Inferior Vena Cava Filter [medd]
637  C1705817: Stent                     [medd]

Meta Mapping (911):
909  C0080306: Inferior Vena Cava Filter [medd]
637  C0038257: Stent                     [medd]
```

Metathesaurus String Growth 1990-2011



U. S. NATIONAL LIBRARY OF MEDICINE

An Especially Egregious Example

- Phrase from PMID 10931555

  protein-4 FN3 fibronectin type III domain GSH glutathione GST glutathione S-transferase hIL-6 human interleukin-6 HSA human serum albumin IC(50) half-maximal inhibitory concentration Ig immunoglobulin IMAC immobilized metal affinity chromatography K(D) equilibrium constant
- Extreme, but not atypical
- MetaMap identifies 99 concepts
- Mappings are subsets of candidates: Up to 2^{99} mappings
- Would require 10^{21} TB of memory!
- Algorithmic Solutions

Solution - Pruning the Candidate Set

Inferior vena caval stent filter

```
909  C0080306: Inferior Vena Cava Filter [medd]
804  C0180860: Filter                    [mnob]
804  C0581406: Filter                    [medd]
804  C1522664: Filter                    [inpr]
804  C1704449: Filter                    [cnce]
804  C1704684: Filter                    [medd]
804  C1875155: FILTER                    [medd]
717  C0521360: Inferior vena caval       [blor]
673  C0042460: Vena caval                [bpoc]
637  C0038257: Stent                     [medd]
637  C1705817: Stent                     [medd]
637  C0447122: Vena                      [bpoc]
```

Results of Algorithmic Improvements

- 2010 MEDLINE baseline: 146 troublesome citations
- Original runtime > 12 hours per citation
- Improved runtime ~ 12.3 seconds per citation
- **350,000%** improvement for problematic citations

Efficiency improvements across MEDLINE baseline:

- 2004 MEDLINE Baseline (12.5M citations): 6 months
- 2012 MEDLINE Baseline (20.5M citations): 8 days

Significant New MetaMap Features

Solutions for problems

- Default output difficult to post-process:
  - XML output
- MetaMap originally developed for literature, not clinical:
  - Wendy Chapman’s NegEx (negation detection)
  - User-Defined Acronyms

Literature: Author-Defined Acronyms

Acronyms often defined by authors in literature:

- *Trimethyl cetyl ammonium pentachlorphenate (TCAP) and fatty acids as antifungal agents.*
- *Reticulo-endothelial immune serum (REIS) in a globulin fraction*
- *The bacteriostatic action of isonicotinic acid hydrazid (INAH) on tubercle bacilli*
- *the interstitial latero-dorsal hypothalamic nucleus (ILDHN) of the female guinea pig*
- *The adrenocorticotropic hormone (ACTH) of the anterior pituitary.*

MetaMap replaces acronyms' short form with their long form

Clinical Text: Undefined Acronyms

Acronyms rarely defined in clinical text:

- *He underwent a **CABG** and **PTCA** in 2008.*
- *EKGs show a **RBBB** with **LAFB** with 1st **AV** block*
- *Sequential **LIMA** to the diagonal and **LAD** and sequential **SVG** to the **PLB** and **PDA** and **SVG** to **IM** grafts were placed*

post-transplantation lymphoproliferative disorder

- ***LAD** and **SVG** to **D1** patent*
- *treatment for **PTLD** with Rituxan versus **CHOP***

MetaMap

cyclophosphamide, hydroxydaunomycin, Oncovin, and prednisone

Allows customizations tailored to specific needs

User-Defined Acronyms (UDAs)

Customize UDAs for radiol

- **CAT | Computerized A**
- **PET | Positron Emiss**

Otherwise.....................

```
C0031268: Pet (Pet Ani
C1456682: Pets (Pet He
C0007450: Cat (Felis c
C0325090: Cat (Felis s
C0524517: Cat (Genus F
C0325089: cats (Family
```

Outline

- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

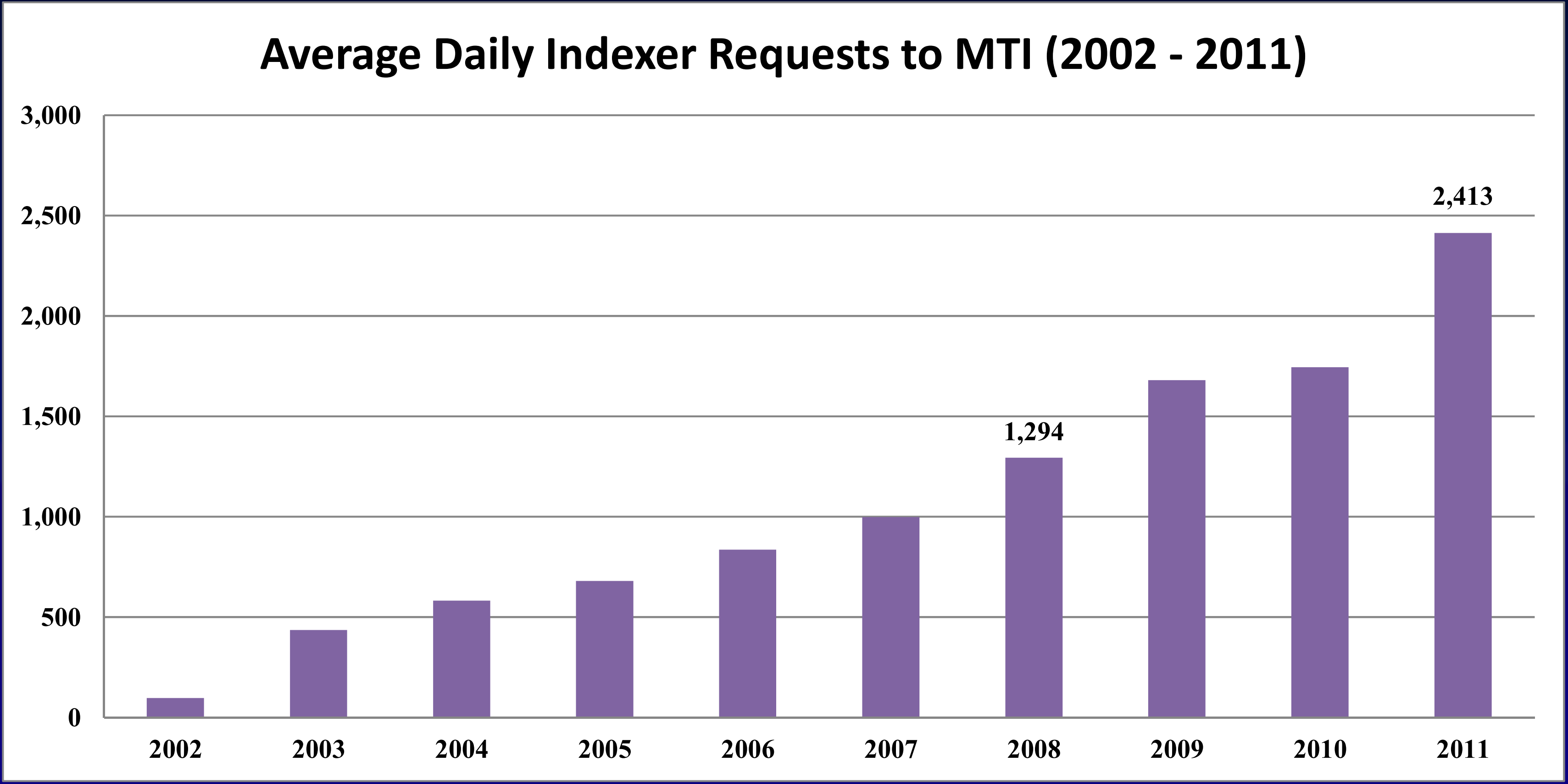
The NLM Medical Text Indexer (MTI)

- Overview
- Uses
- MTI as First-Line Indexer (MTIFL)
- Performance

MTI - Overview

- Summarizes input text into an ordered list of MeSH Headings
- In use since mid-2002
- Developed with continued Index Section collaboration
- Uses article Title and Abstract
- Provides recommendations for 96% of indexed articles
- Indexer consulted for 50% of indexed articles

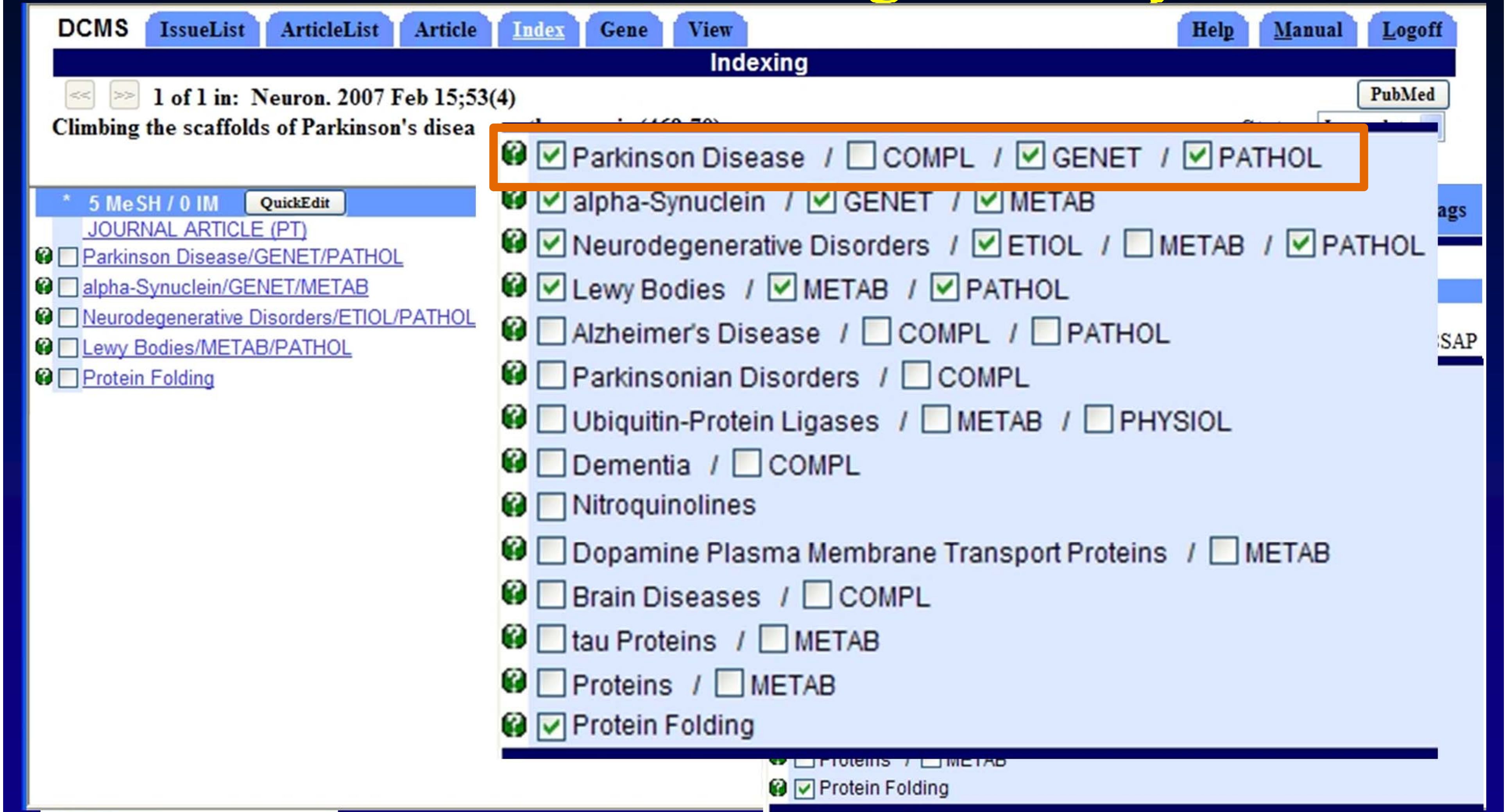
MTI Usage


Average Daily Indexer Requests to MTI (2002 - 2011)
3,000
2,500
2,000
1,500
1,000
500
0
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
1,294
2,413

MTI - Uses

- Assisted indexing of Index Section journal articles
- Assisted indexing of Cataloging and History of Medicine Division records
- Automatic indexing of NLM Gateway meeting abstracts
- First-line indexing (MTIFL) since February 2011
- Also available to the Community
  - 45,000 requests (2011)

Data Creation and Management System



Indexing

<< >> 1 of 1 in: Neuron. 2007 Feb 15;53(4)

Climbing the scaffolds of Parkinson's disea

* 5 MeSH / 0 IM

JOURNAL ARTICLE (PT)
- □ Parkinson Disease/GENET/PATHOL
- □ alpha-Synuclein/GENET/METAB
- □ Neurodegenerative Disorders/ETIOL/PATHOL
- □ Lewy Bodies/METAB/PATHOL
- □ Protein Folding

- ☑ Parkinson Disease / □ COMPL / ☑ GENET / ☑ PATHOL
- ☑ alpha-Synuclein / ☑ GENET / ☑ METAB
- ☑ Neurodegenerative Disorders / ☑ ETIOL / □ METAB / ☑ PATHOL
- ☑ Lewy Bodies / ☑ METAB / ☑ PATHOL
- □ Alzheimer's Disease / □ COMPL / □ PATHOL
- □ Parkinsonian Disorders / □ COMPL
- □ Ubiquitin-Protein Ligases / □ METAB / □ PHYSIOL
- □ Dementia / □ COMPL
- □ Nitroquinolines
- □ Dopamine Plasma Membrane Transport Proteins / □ METAB
- □ Brain Diseases / □ COMPL
- □ tau Proteins / □ METAB
- □ Proteins / □ METAB
- ☑ Protein Folding

ags

SAP

- □ Proteins / □ METAB
- ☑ Protein Folding

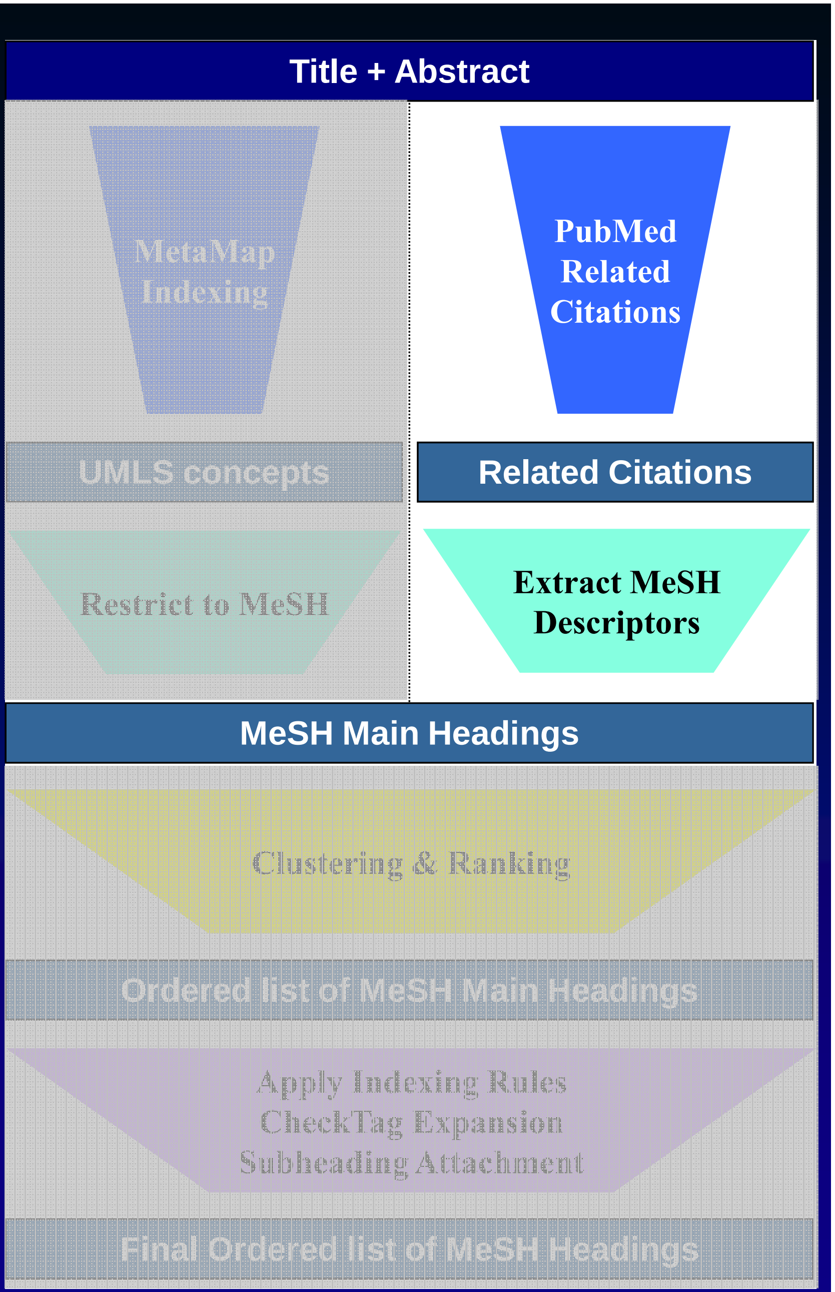
U. S. NATIONAL LIBRARY OF MEDICINE

MTI - Uses

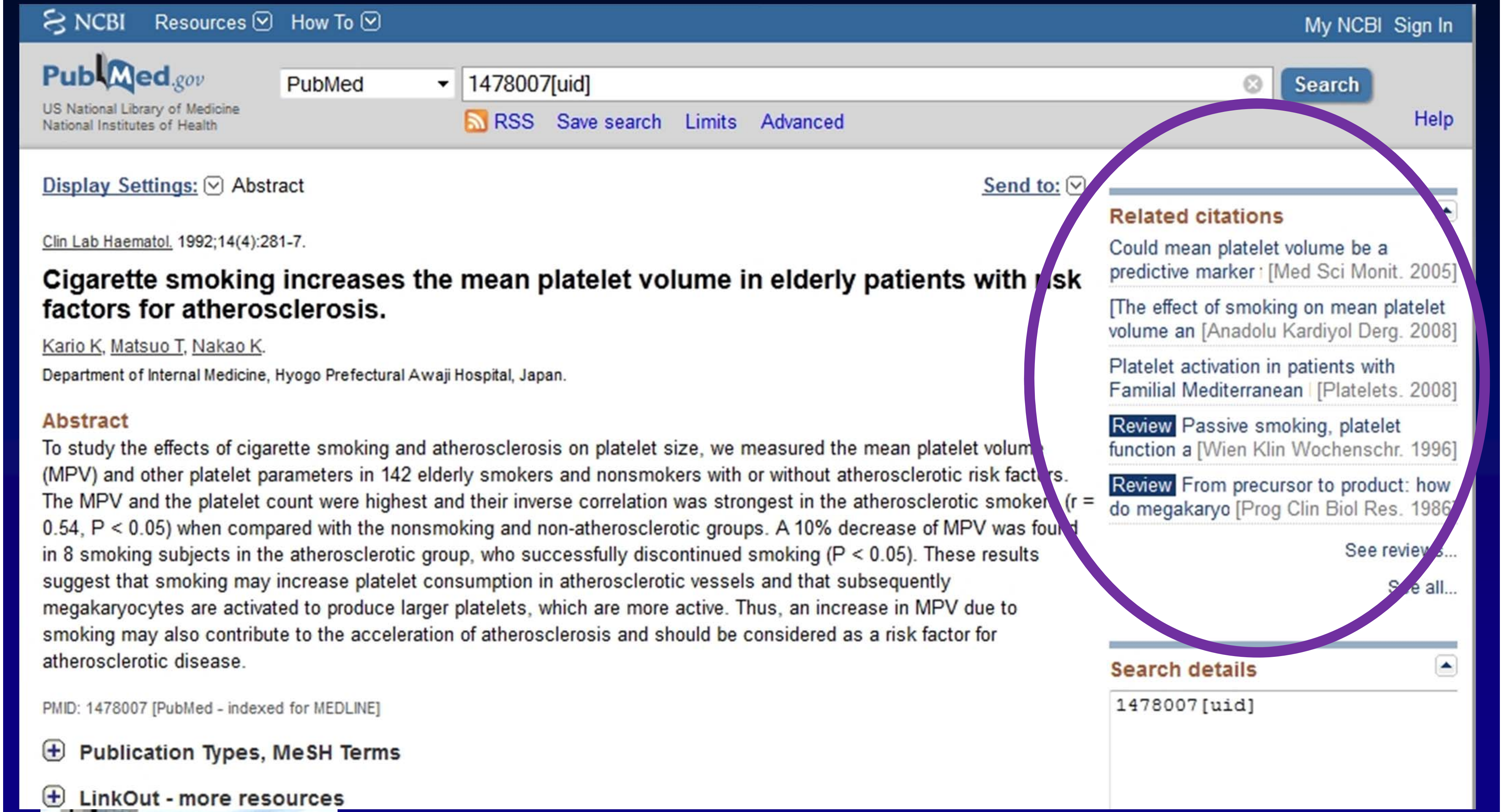
- Assisted indexing of Index Section journal articles
- Assisted indexing of Cataloging and History of Medicine Division records
- Automatic indexing of NLM Gateway meeting abstracts
- First-line indexing (MTIFL) since February 2011
- Also available to the Community
  - 45,000 requests (2011)

MTI

- MetaMap Indexing – Actually found in text
- Restrict to MeSH – Maps UMLS Concepts to MeSH
- PubMed Related Citations – Not necessarily found in text



PubMed Query Example



1478007[uid]

Display Settings: Abstract

Clin Lab Haematol. 1992;14(4):281-7.

Cigarette smoking increases the mean platelet volume in elderly patients with risk factors for atherosclerosis.

Kario K, Matsuo T, Nakao K.

Department of Internal Medicine, Hyogo Prefectural Awaji Hospital, Japan.

Abstract

To study the effects of cigarette smoking and atherosclerosis on platelet size, we measured the mean platelet volume (MPV) and other platelet parameters in 142 elderly smokers and nonsmokers with or without atherosclerotic risk factors. The MPV and the platelet count were highest and their inverse correlation was strongest in the atherosclerotic smokers (r = 0.54, P < 0.05) when compared with the nonsmoking and non-atherosclerotic groups. A 10% decrease of MPV was found in 8 smoking subjects in the atherosclerotic group, who successfully discontinued smoking (P < 0.05). These results suggest that smoking may increase platelet consumption in atherosclerotic vessels and that subsequently megakaryocytes are activated to produce larger platelets, which are more active. Thus, an increase in MPV due to smoking may also contribute to the acceleration of atherosclerosis and should be considered as a risk factor for atherosclerotic disease.

PMID: 1478007 [PubMed - indexed for MEDLINE]

⊕ **Publication Types, MeSH Terms**

⊕ **LinkOut - more resources**

Related citations

Could mean platelet volume be a predictive marker [Med Sci Monit. 2005]

[The effect of smoking on mean platelet volume an [Anadolu Kardiyol Derg. 2008]

Platelet activation in patients with Familial Mediterranean [Platelets. 2008]

Review Passive smoking, platelet function a [Wien Klin Wochenschr. 1996]

Review From precursor to product: how do megakaryo [Prog Clin Biol Res. 1986]

See reviews...

See all...

Search details

1478007[uid]

MTI

- MetaMap Indexing – Actually found in text

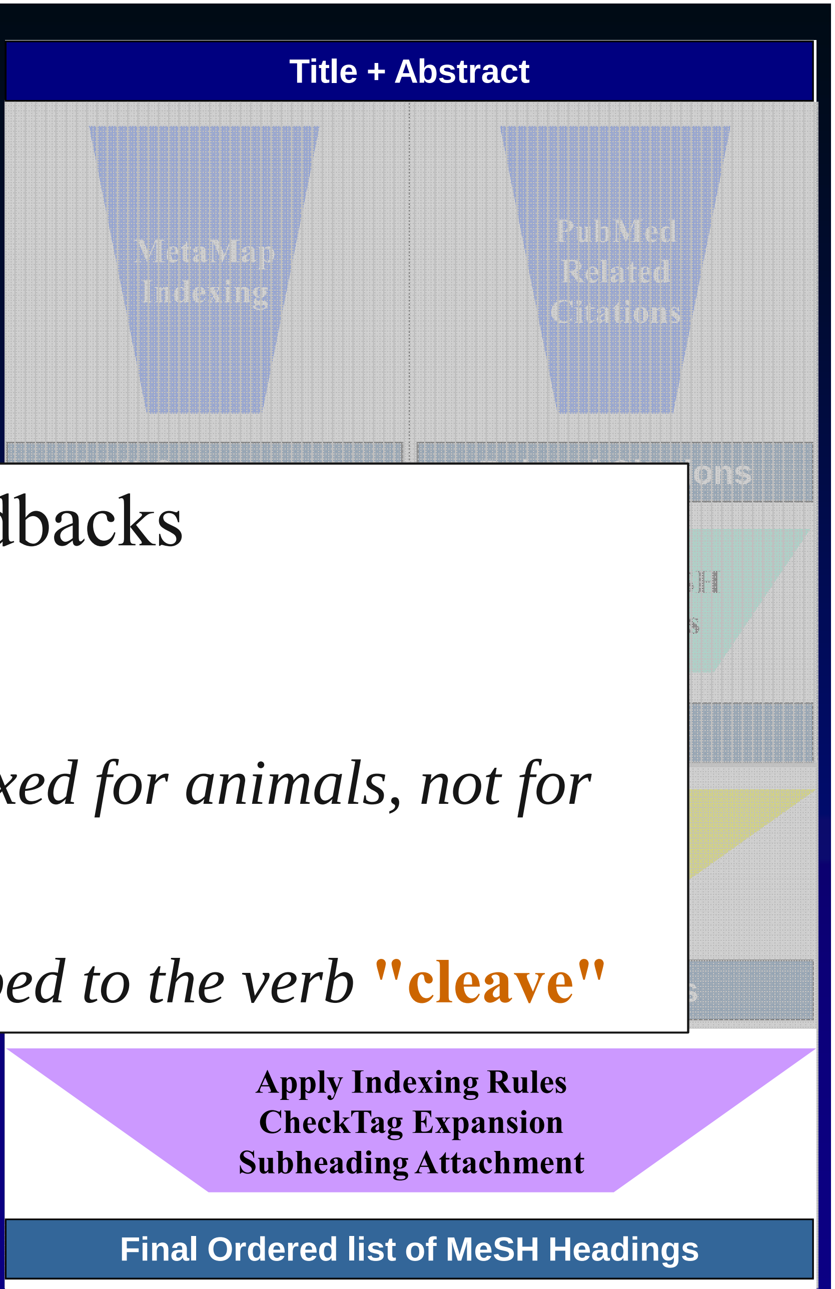
Received **2,330** Indexer Feedbacks

Incorporated **40%** into MTI

March 20, 2012

Hibernation *should only be indexed for animals, not for* **"stem cell hibernation"**

Clove *(spice) should not be mapped to the verb* **"cleave"**



Title + Abstract

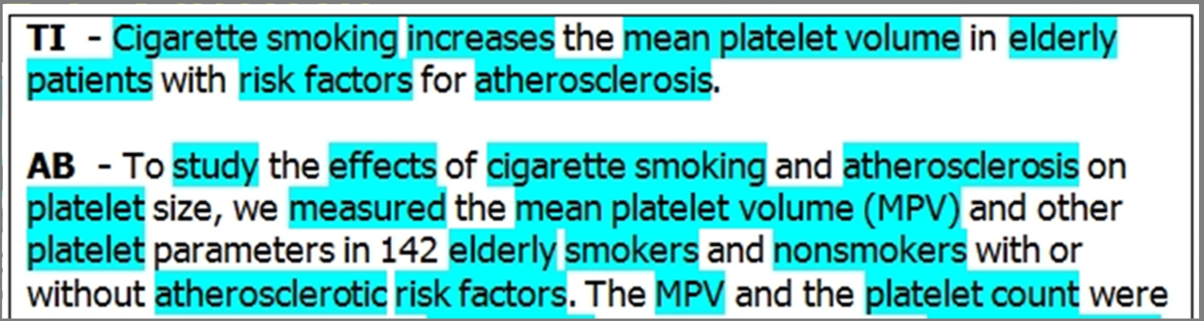
MetaMap Indexing

PubMed Related Citations

Apply Indexing Rules
CheckTag Expansion
Subheading Attachment

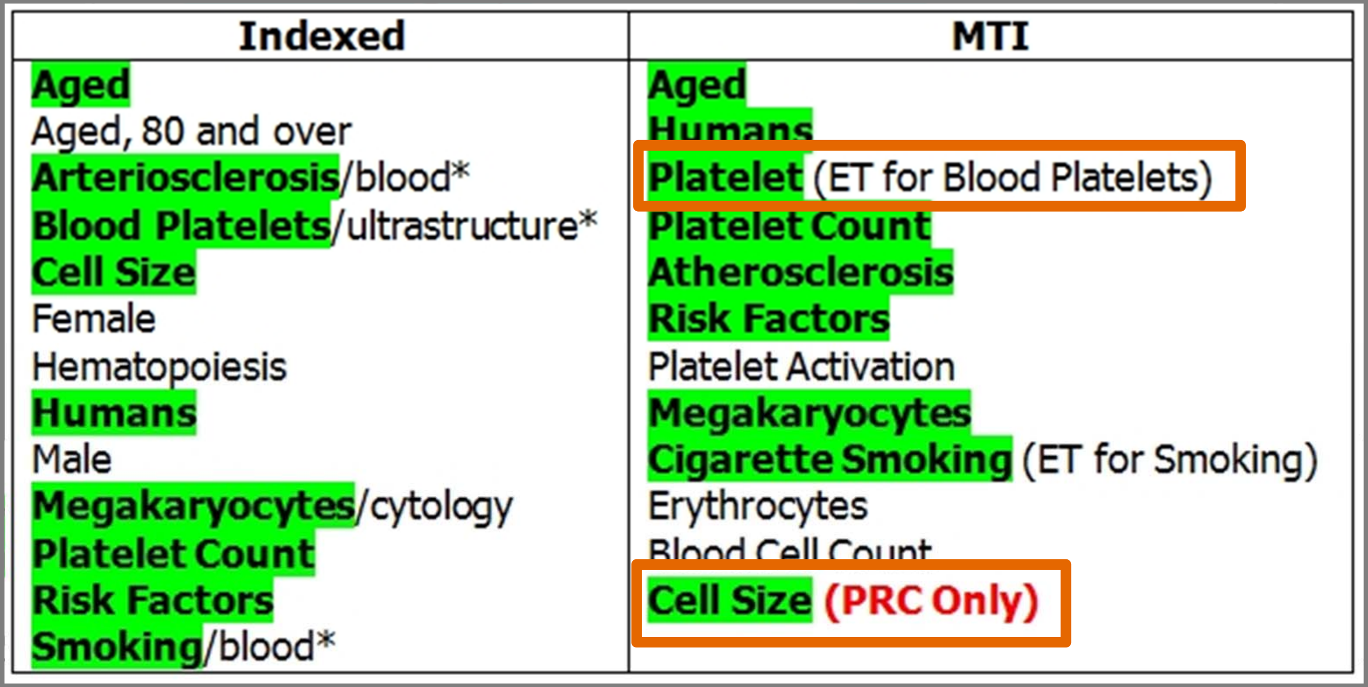
Final Ordered list of MeSH Headings

M



TI - Cigarette smoking increases the mean platelet volume in elderly patients with risk factors for atherosclerosis.

AB - To study the effects of cigarette smoking and atherosclerosis on platelet size, we measured the mean platelet volume (MPV) and other platelet parameters in 142 elderly smokers and nonsmokers with or without atherosclerotic risk factors. The MPV and the platelet count were

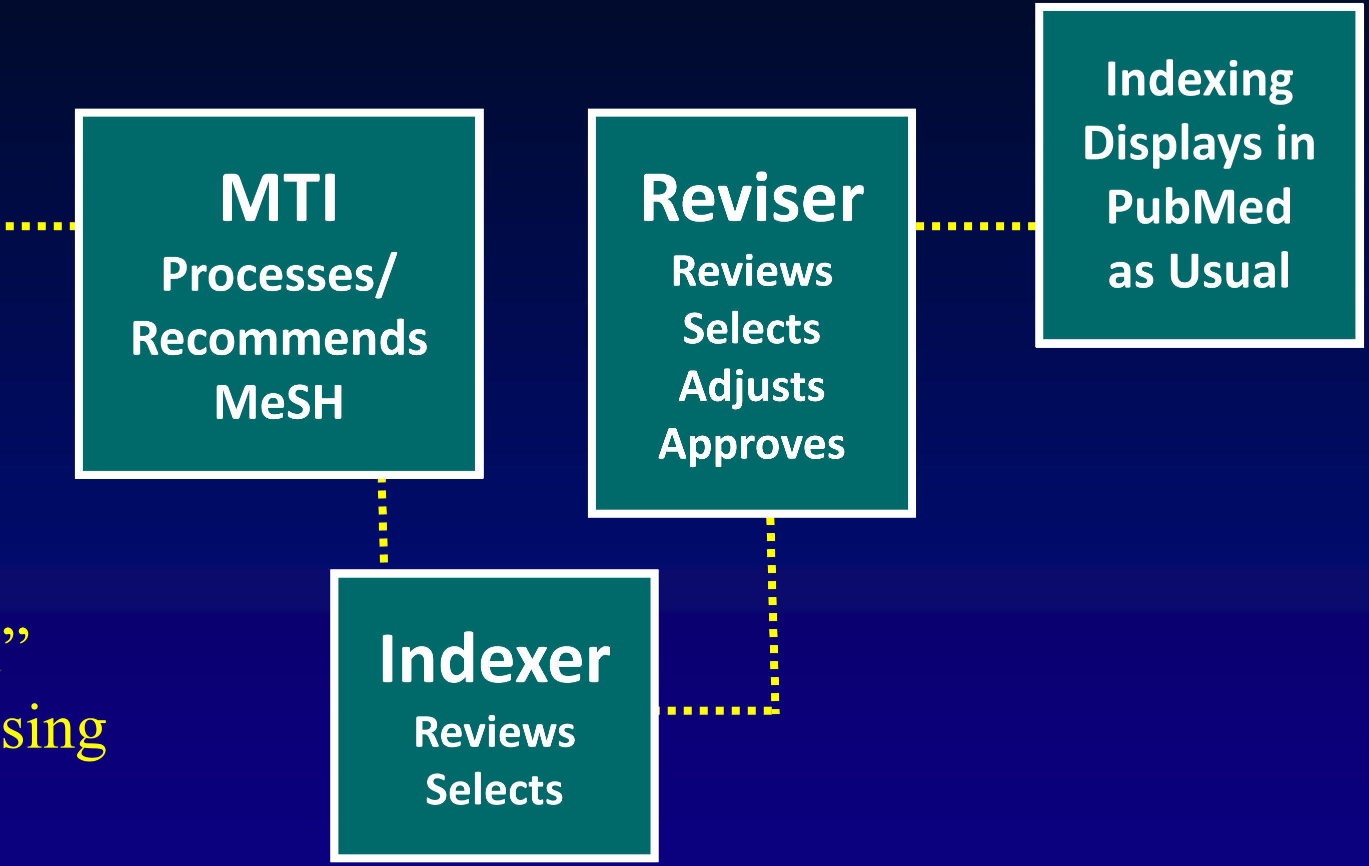


Indexed	MTI
Aged	**Aged**
Aged, 80 and over	**Humans**
Arteriosclerosis/blood*	**Platelet** (ET for Blood Platelets)
Blood Platelets/ultrastructure*	**Platelet Count**
Cell Size	**Atherosclerosis**
Female	**Risk Factors**
Hematopoiesis	Platelet Activation
Humans	**Megakaryocytes**
Male	**Cigarette Smoking** (ET for Smoking)
Megakaryocytes/cytology	Erythrocytes
Platelet Count	Blood Cell Count
Risk Factors	**Cell Size** **(PRC Only)**
Smoking/blood*	

PMID: 147

MTI as First-Line Indexer (MTIFL)

23 MEDLINE Journals



MTI
Processes/ Recommends MeSH

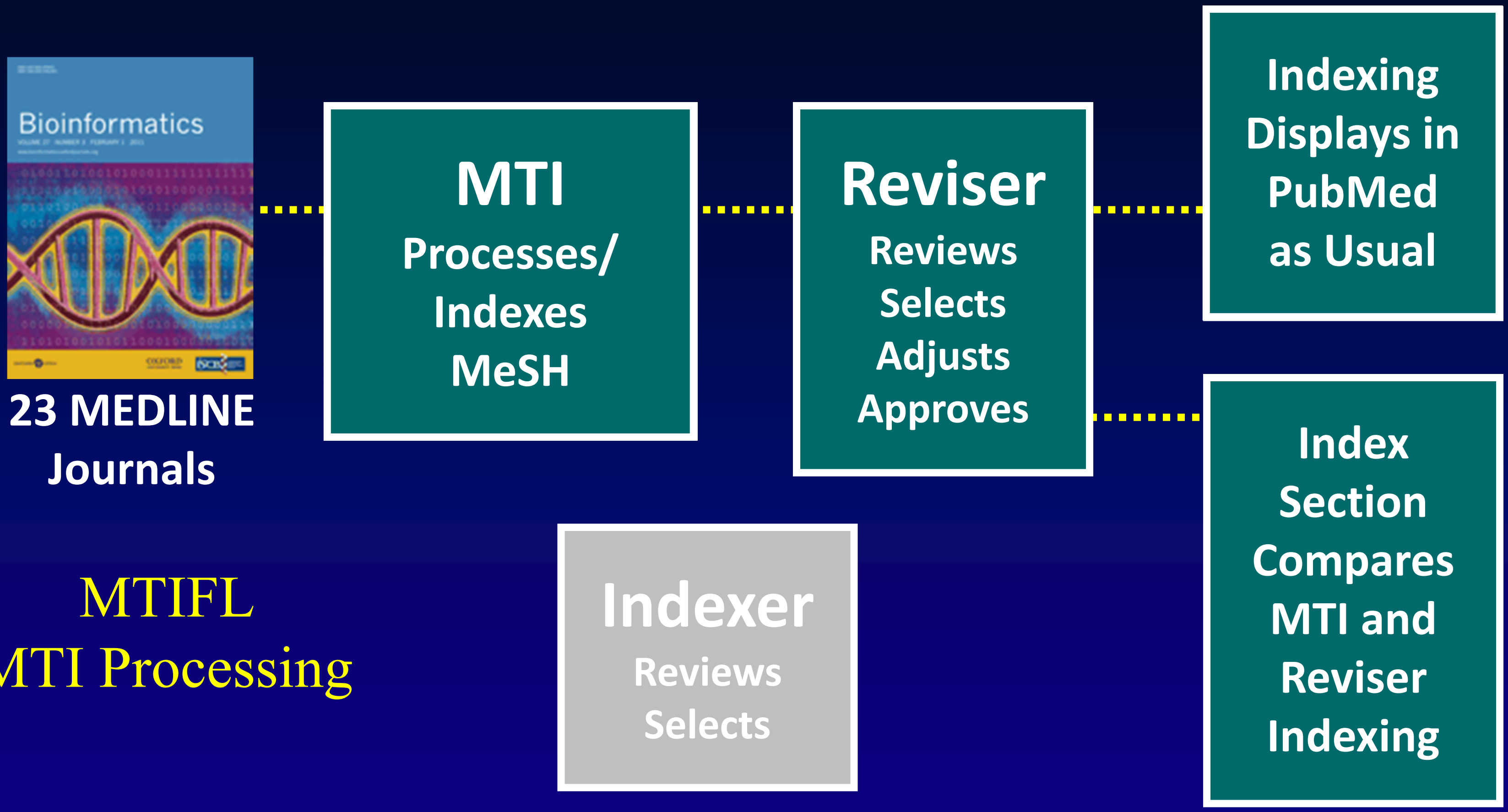
Reviser
Reviews
Selects
Adjusts
Approves

Indexing Displays in PubMed as Usual

Indexer
Reviews
Selects

“Normal” MTI Processing

MTI as First-Line Indexer (MTIFL)



23 MEDLINE Journals

MTI
Processes/
Indexes
MeSH

Reviser
Reviews
Selects
Adjusts
Approves

Indexing Displays in PubMed as Usual

Index Section Compares MTI and Reviser Indexing

Indexer
Reviews
Selects

MTIFL
MTI Processing

MTIFL

- Experiments in 2010 led by Marina Rappaport
  - Microbiology, Anatomy, Botany, and Medical Informatics journals
- Initial experiment involved both Indexers and MTI
  - Provided baseline timings and performance

	Indexer	MTIFL	Diff
Number of Articles	609	668	
Average Total Minutes	12.05	14.37	+2.32
Average MHs	11.12	24.75	+13.63

  - Identified challenges (and opportunities)
    - Publication Types
    - Chemical Flags
    - Functional annotation of genes

} Manually added by indexer

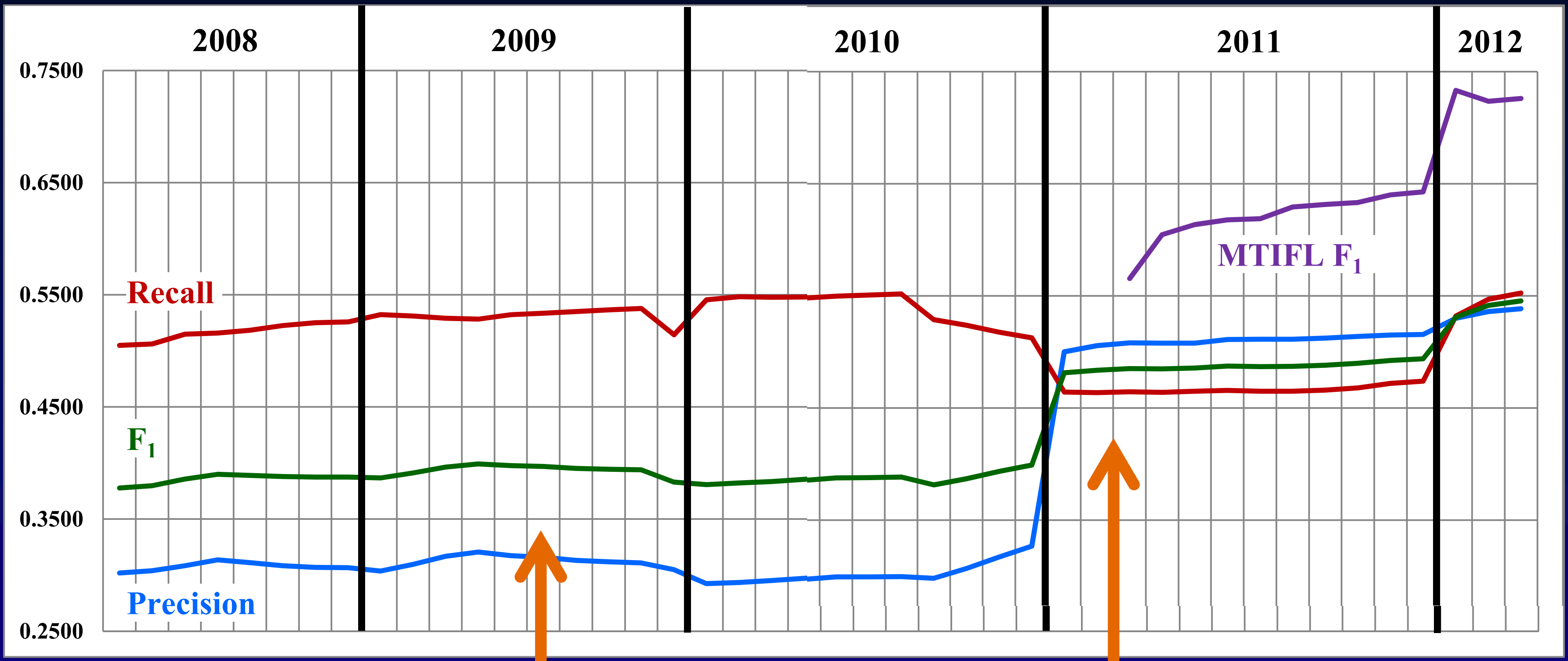
MTIFL

- Follow-on experiments focused on reducing MTI revision time:
  - Reduce the number of MTI indexing terms
  - Focus on journals with few/no Gene Annotation or Chemical Flags

	Initial		Final	
	Indexer	MTIFL	MTIFL	MTIFL Diff
Average Total Minutes	12.05	14.37	10.01	**-4.36**
Average MHs	11.12	24.75	8.58	**-16.17**

- MTI revision time **2.04 minutes faster** than Indexer revised time (10.01 minutes vs 12.05 minutes)
- Pilot project started with 14 journals, expanded to 23 in 2011

MTI - How are we doing?



Fruition of 2011 Changes

Focus on Precision versus Recall

Outline

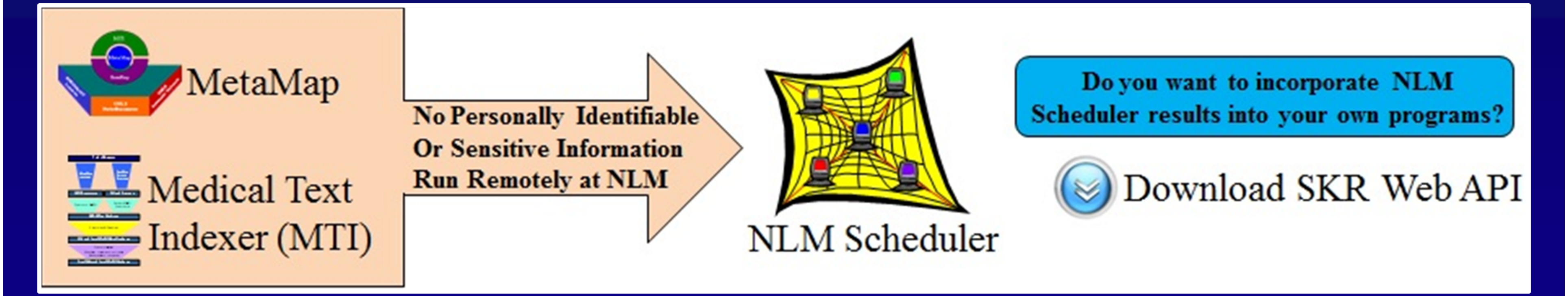
- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

Availability of Indexing Initiative Tools

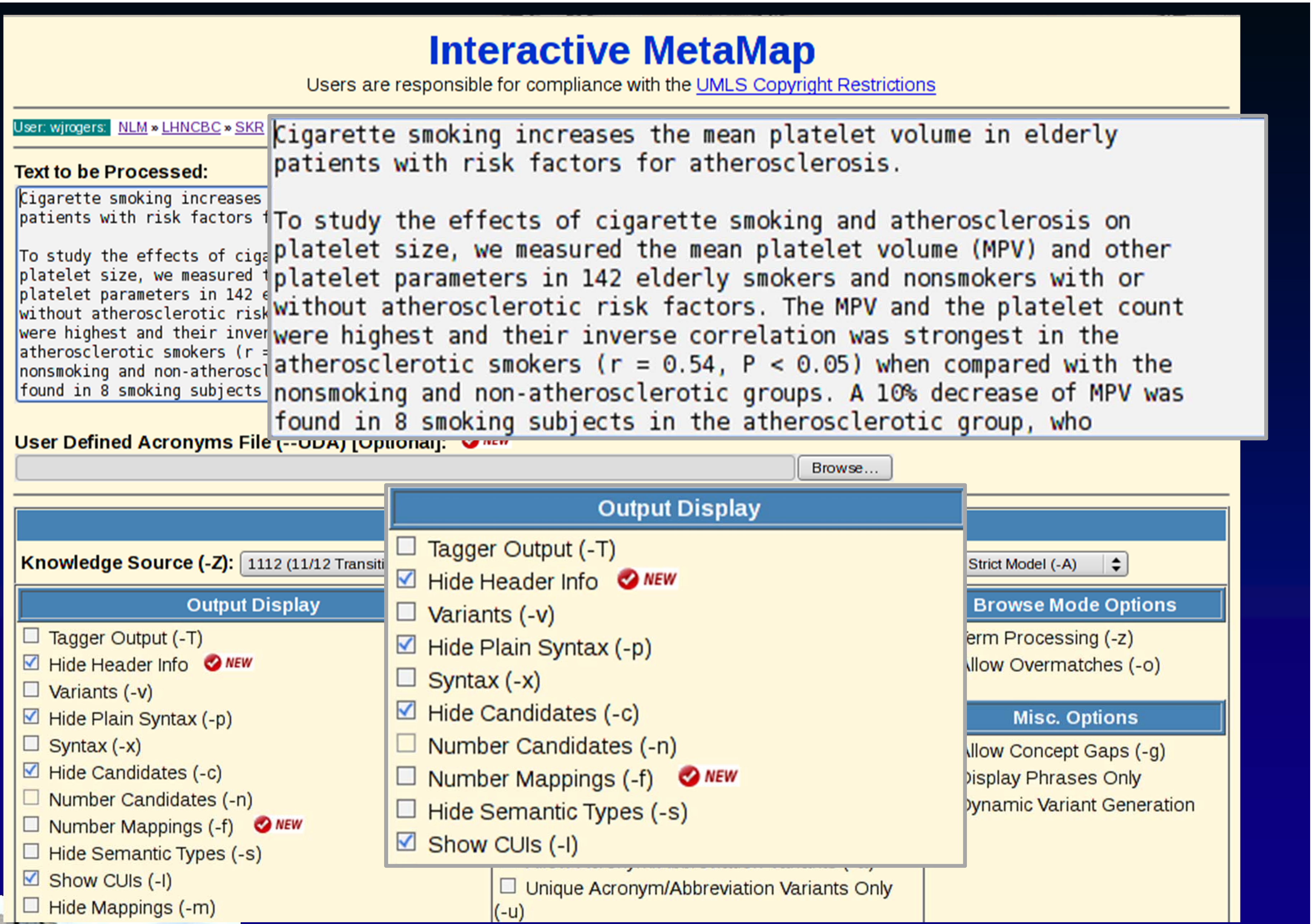
- Remote Access
  - Web
  - API
- Local Installation
  - Linux
  - Mac OS/X
  - Windows XP/7

Remote Access

- Interactive
  - Small input data (for testing, etc.), immediate results
- Batch
  - Large input data processed using a large pool of computing resources


MetaMap
Medical Text Indexer (MTI)
No Personally Identifiable Or Sensitive Information Run Remotely at NLM
NLM Scheduler
Do you want to incorporate NLM Scheduler results into your own programs?
Download SKR Web API

Interactive MetaMap



Users are responsible for compliance with the UMLS Copyright Restrictions

User: wjrogers: NLM » LHNCBC » SKR

Text to be Processed:

```
Cigarette smoking increases the mean platelet volume in elderly
patients with risk factors for atherosclerosis.

To study the effects of cigarette smoking and atherosclerosis on
platelet size, we measured the mean platelet volume (MPV) and other
platelet parameters in 142 elderly smokers and nonsmokers with or
without atherosclerotic risk factors. The MPV and the platelet count
were highest and their inverse correlation was strongest in the
atherosclerotic smokers (r = 0.54, P < 0.05) when compared with the
nonsmoking and non-atherosclerotic groups. A 10% decrease of MPV was
found in 8 smoking subjects in the atherosclerotic group, who
```

```
Cigarette smoking increases
patients with risk factors f

To study the effects of ciga
platelet size, we measured t
platelet parameters in 142 e
without atherosclerotic risk
were highest and their inver
atherosclerotic smokers (r =
nonsmoking and non-atherosc]
found in 8 smoking subjects
```

User Defined Acronyms File (--UDA) [Optional]: NEW

Browse...

Knowledge Source (-Z): 1112 (11/12 Transiti

Strict Model (-A)

Output Display

- □ Tagger Output (-T)
- ☑ Hide Header Info NEW
- □ Variants (-v)
- ☑ Hide Plain Syntax (-p)
- □ Syntax (-x)
- ☑ Hide Candidates (-c)
- □ Number Candidates (-n)
- □ Number Mappings (-f) NEW
- □ Hide Semantic Types (-s)
- ☑ Show CUIs (-I)
- □ Hide Mappings (-m)

Output Display

- □ Tagger Output (-T)
- ☑ Hide Header Info NEW
- □ Variants (-v)
- ☑ Hide Plain Syntax (-p)
- □ Syntax (-x)
- ☑ Hide Candidates (-c)
- □ Number Candidates (-n)
- □ Number Mappings (-f) NEW
- □ Hide Semantic Types (-s)
- ☑ Show CUIs (-I)

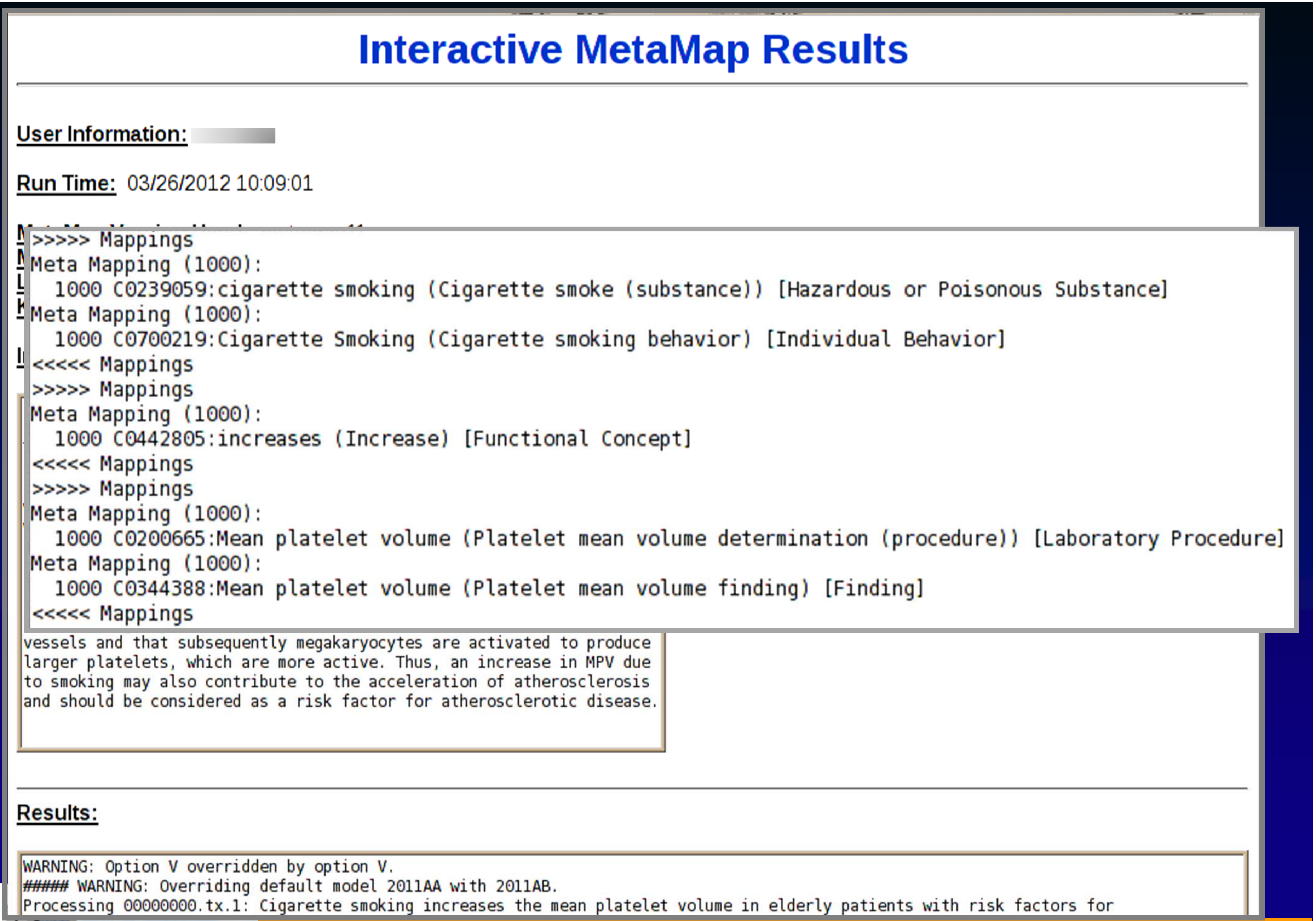
- □ Unique Acronym/Abbreviation Variants Only (-u)

Browse Mode Options

erm Processing (-z)
llow Overmatches (-o)

Misc. Options

llow Concept Gaps (-g)
isplay Phrases Only
ynamic Variant Generation



Interactive MetaMap Results

User Information:

Run Time: 03/26/2012 10:09:01

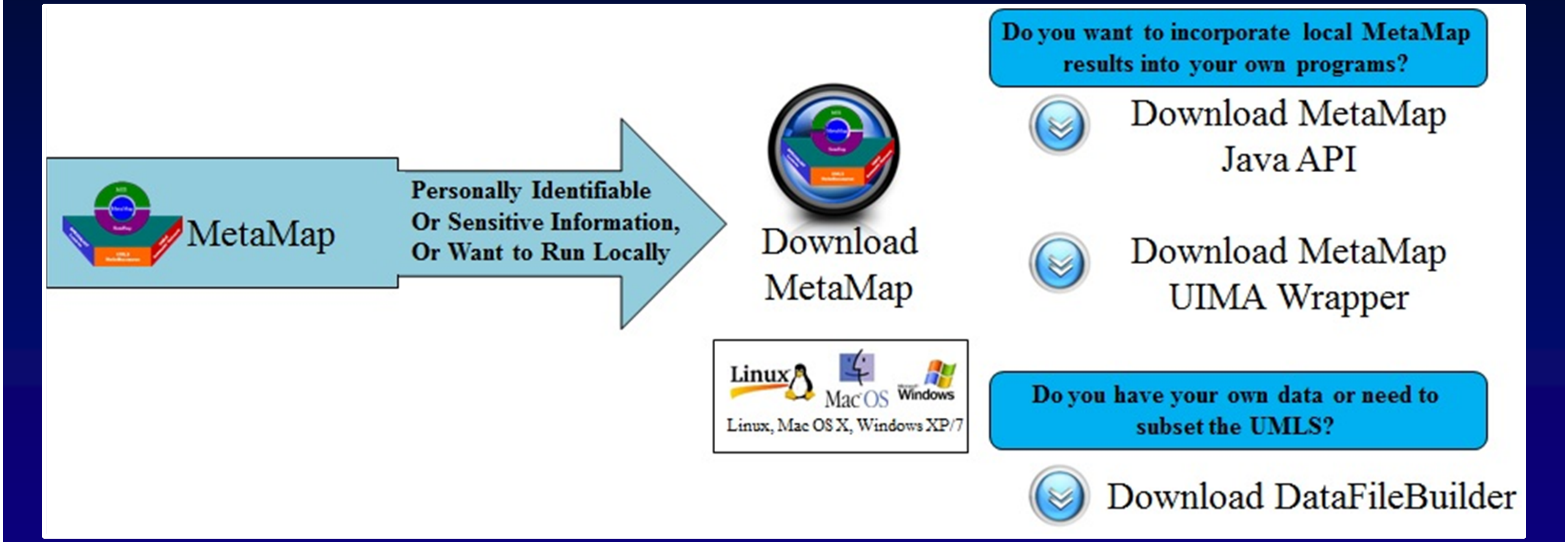
```
>>>>> Mappings
Meta Mapping (1000):
  1000 C0239059:cigarette smoking (Cigarette smoke (substance)) [Hazardous or Poisonous Substance]
Meta Mapping (1000):
  1000 C0700219:Cigarette Smoking (Cigarette smoking behavior) [Individual Behavior]
<<<<< Mappings
>>>>> Mappings
Meta Mapping (1000):
  1000 C0442805:increases (Increase) [Functional Concept]
<<<<< Mappings
>>>>> Mappings
Meta Mapping (1000):
  1000 C0200665:Mean platelet volume (Platelet mean volume determination (procedure)) [Laboratory Procedure]
Meta Mapping (1000):
  1000 C0344388:Mean platelet volume (Platelet mean volume finding) [Finding]
<<<<< Mappings
```

```
vessels and that subsequently megakaryocytes are activated to produce
larger platelets, which are more active. Thus, an increase in MPV due
to smoking may also contribute to the acceleration of atherosclerosis
and should be considered as a risk factor for atherosclerotic disease.
```

Results:

```
WARNING: Option V overridden by option V.
###### WARNING: Overriding default model 2011AA with 2011AB.
Processing 00000000.tx.1: Cigarette smoking increases the mean platelet volume in elderly patients with risk factors for
```

Local Installation of MetaMap



MetaMap

Personally Identifiable
Or Sensitive Information,
Or Want to Run Locally

Download
MetaMap

Linux, Mac OS X, Windows XP/7

Do you want to incorporate local MetaMap results into your own programs?

Download MetaMap
Java API

Download MetaMap
UIMA Wrapper

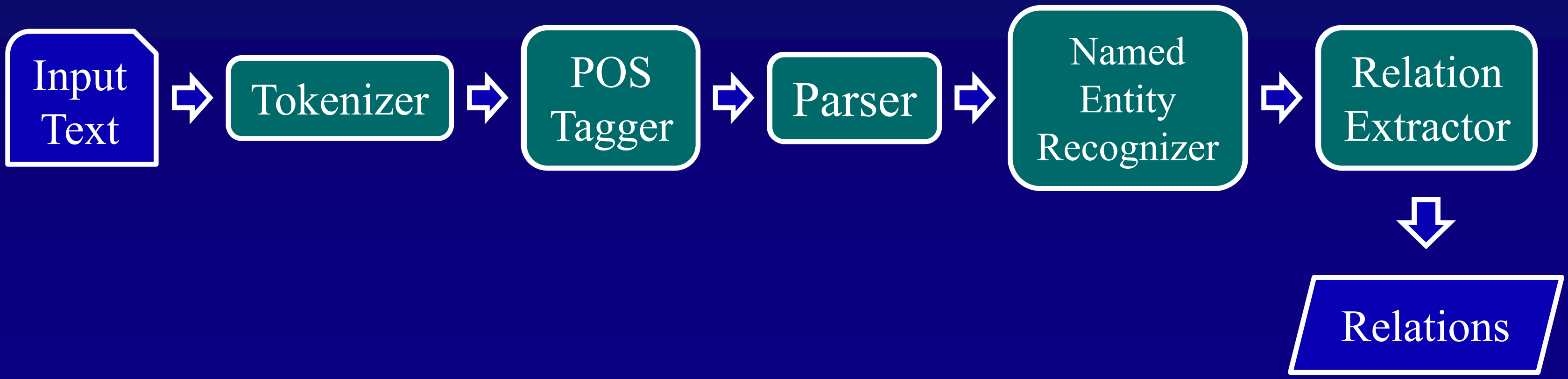
Do you have your own data or need to subset the UMLS?

Download DataFileBuilder

MetaMap as a UIMA Component

- Allows MetaMap to be used as an UIMA “annotator” component.
- UIMA - Unstructured Information Management Architecture

a component-based software for the analysis of unstructured information.



Input Text

Tokenizer

POS Tagger

Parser

Named Entity Recognizer

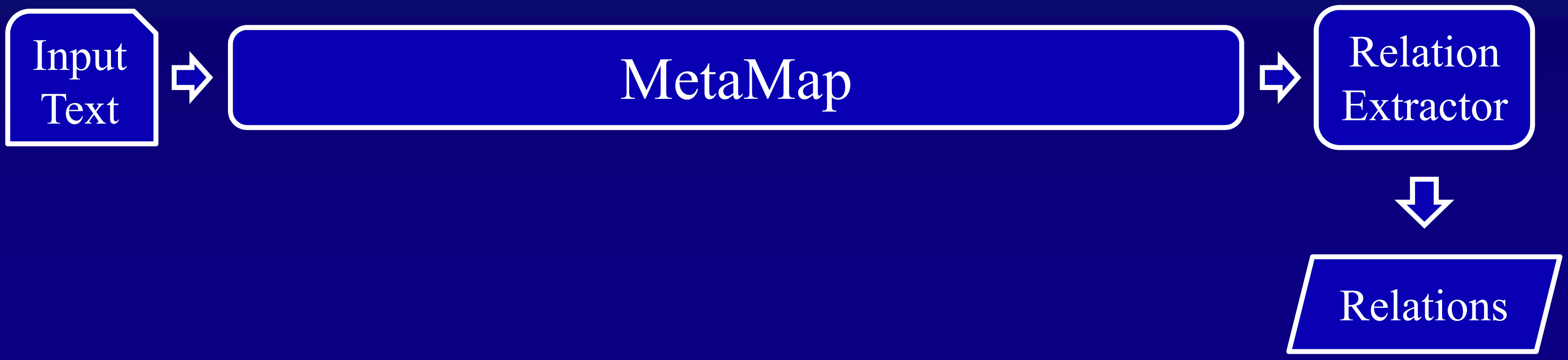
Relation Extractor

Relations

MetaMap as a UIMA Component

- Allows MetaMap to be used as an UIMA “annotator” component.
- UIMA - Unstructured Information Management Architecture

a component-based software for the analysis of unstructured information.



UIMA-compliant NLP Toolkits

- A number of NLP toolkits that are UIMA compliant
  - OpenNLP
  - clinical Text Analysis and Knowledge Extraction System (cTAKES)
  - OpenPipeline

Data File Builder

Provides the ability to create specialized data models for MetaMap:

- UMLS augmented with user data
- UMLS subsets
- Independent knowledge sources
  - Should have notion of concept, synonymy
  - Ontologies
  - Local Thesauri
  - Other Knowledge Sources

Web Access Statistics (2011)

- Remote Access:
  - 7,500 unique visits - 124 different countries
  - 70,000 Interactive Requests
  - 87,000 Batch Requests
- MetaMap Downloads:
  - 1,050 for MetaMap program
    - 570 Linux, 200 Mac/OS, 280 Windows
  - 41 for Data File Builder

Outline

- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

Enhancing MetaMap and MTI Performance

- MetaMap precision enhancement through knowledge-based Word Sense Disambiguation

- MTI enhancement based on Machine Learning

Word Sense Disambiguation (WSD)

- Kids with *colds* may also have a sore throat, cough, headache, mild fever, fatigue, muscle aches, and loss of appetite.

- Candidate MetaMap mappings for *cold*

```
C0234192: Cold (Cold sensation)
C0009264: Cold (Cold temperature)
C0009443: Cold (Common cold)
```

Knowledge-based WSD

- Compare UMLS candidate concept profile vectors to context of ambiguous word
- Concept profile vectors' words from definition, synonyms and related concepts

Common cold	
Weight	**Word**
265	infect
126	disease
41	fever
40	cough

Cold temperature	
Weight	**Word**
258	temperature
86	hypothermia
72	effect
48	hot

- Candidate concept with highest similarity is predicted

Knowledge-based WSD

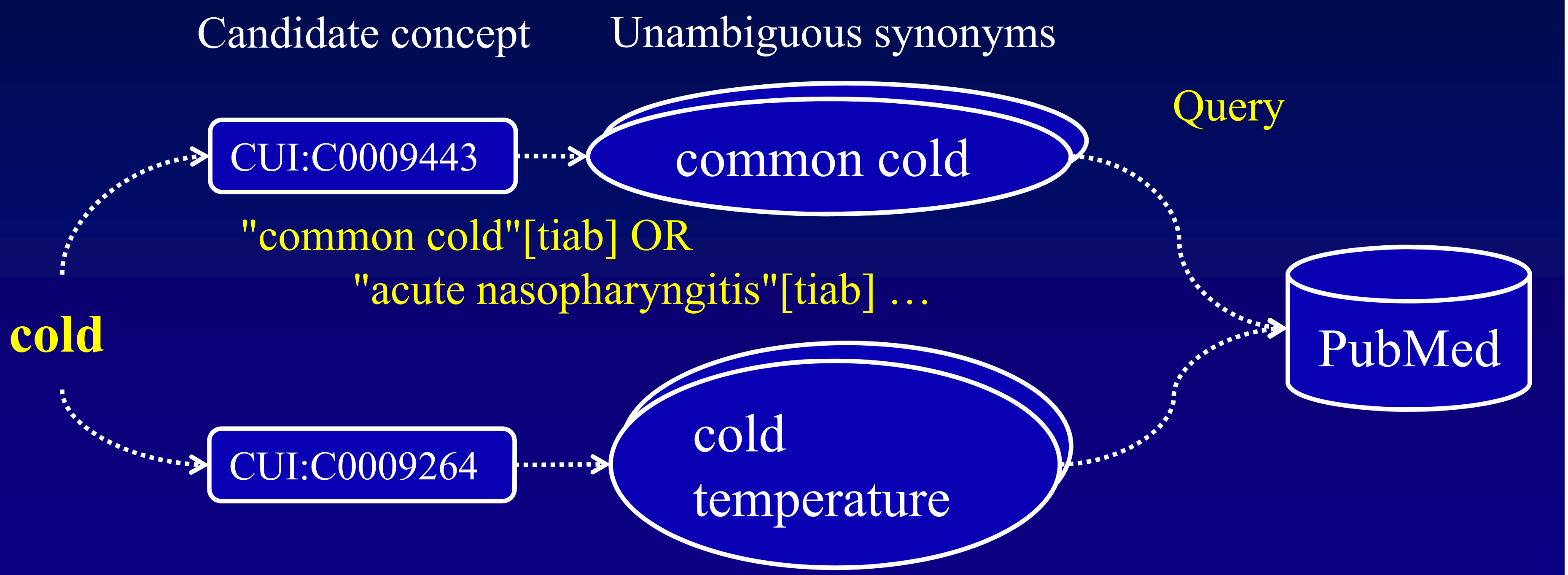
- Kids with *colds* may also have a sore throat, *cough*, headache, mild *fever*, fatigue, muscle aches, and loss of appetite.

Common cold	
Weight	**Word**
265	infect
126	disease
41	**fever**
40	**cough**

Cold temperature	
Weight	**Word**
258	temperature
86	hypothermia
72	effect
48	hot

Automatically Extracted Corpus WSD

- MEDLINE contains numerous examples of ambiguous words context, though not disambiguated



WSD Method Results

- Corpus method has better accuracy than UMLS method

	UMLS	Corpus
NLM WSD	0.65	**0.69**
MSH WSD	0.81	**0.84**

- MSH WSD data set created using MeSH indexing
  - 203 ambiguous words
  - 81 semantic types
  - 37,888 ambiguity cases
- Indirect evaluation with summarization and MTI correlates with direct evaluation

Citation indexed w/Female, Humans and Male

TI -Documenting the symptom experience of cancer patients

AB - Cancer patients experience symptoms associated with their disease, treatment, and comorbidities. Symptom experience is complicated, reflecting symptom prevalence, frequency, and severity. Symptom burden is associated with treatment tolerance as well as patients' quality of life (QOL). A convenience sample of patients with the five most common cancers at a comprehensive cancer center completed surveys assessing symptom experience (Memorial Symptom Assessment Survey) and QOL (Functional Assessment of Cancer Therapy). Patients completed surveys at baseline and at 3, 6, 9, and 12 months thereafter. Surveys were completed by 558 cancer patients with breast, colorectal, gynecologic, lung, or prostate cancer. Patients reported an average of 9.1 symptoms, with symptom experience varying by cancer type. The mean overall QOL for the total sample was 85.1, with results differing by cancer type. Prostate cancer patients reported the lowest symptom burden and the highest QOL. The symptom experience of cancer patients varies widely depending on cancer type. Nevertheless, most patients report symptoms, regardless of whether or not they are currently receiving treatment.

MTI enhancement with Machine Learning

- Large number of indexing examples available from MEDLINE
- Two approaches
  - Semi-automatic generation of indexing rules
  - Indexing algorithm selection through meta-learning

Bottom-up Indexing Approach

- Automatic analysis of citations
  - selection of terms
  - production of candidate annotation rules
- Manual examination and processing
- Post-filtering based on machine learning
- Works well with some MeSH headings; e.g. ‘Carbohydrate Sequence’

MTI Meta-Learning

- No single method performs better than all evaluated indexing methods
- Manual selection of best performing indexing methods becomes tedious with a large number of MHs
- Select indexing methods automatically based on meta-learning

CheckTags Machine Learning Results

- 200k citations for training and 100k citations for testing

CheckTag	F_1 before ML	F_1 with ML	Improvement
Middle Aged	1.01%	59.50%	+58.49
Aged	11.72%	54.67%	+42.95
Child, Preschool	6.11%	45.40%	+39.29
Adult	19.49%	56.84%	+37.35
Male	38.47%	71.14%	+32.67
Aged, 80 and over	1.50%	30.89%	+29.39
Young Adult	2.83%	31.63%	+28.80
Female	46.06%	73.84%	+27.78
Adolescent	24.75%	42.36%	+17.61
Humans	79.98%	91.33%	+11.35
Infant	34.39%	44.69%	+10.30
Swine	71.04%	74.75%	+3.71

CheckTags Machine Learning Results

- 200k citations for training and 100k citations for testing

CheckTag	**F$_1$ before ML**	**F$_1$ with ML**	**Improvement**
Middle Aged	1.01%	59.50%	+58.49
Aged	11.72%	54.67%	+42.95
Child, Preschool	6.11%	45.40%	+39.29
Adult	19.49%	56.84%	+37.35
Male	**38.47%**	71.14%	+32.67
Aged, 80 and over	1.50%	30.89%	+29.39
Young Adult	2.83%	31.63%	+28.80
Female	**46.06%**	**73.84%**	**+27.78**
Adolescent	24.75%	42.36%	+17.61
Humans	**79.98%**	**91.33%**	**+11.35**
Infant	34.39%	44.69%	+10.30
Swine	71.04%	74.75%	+3.71

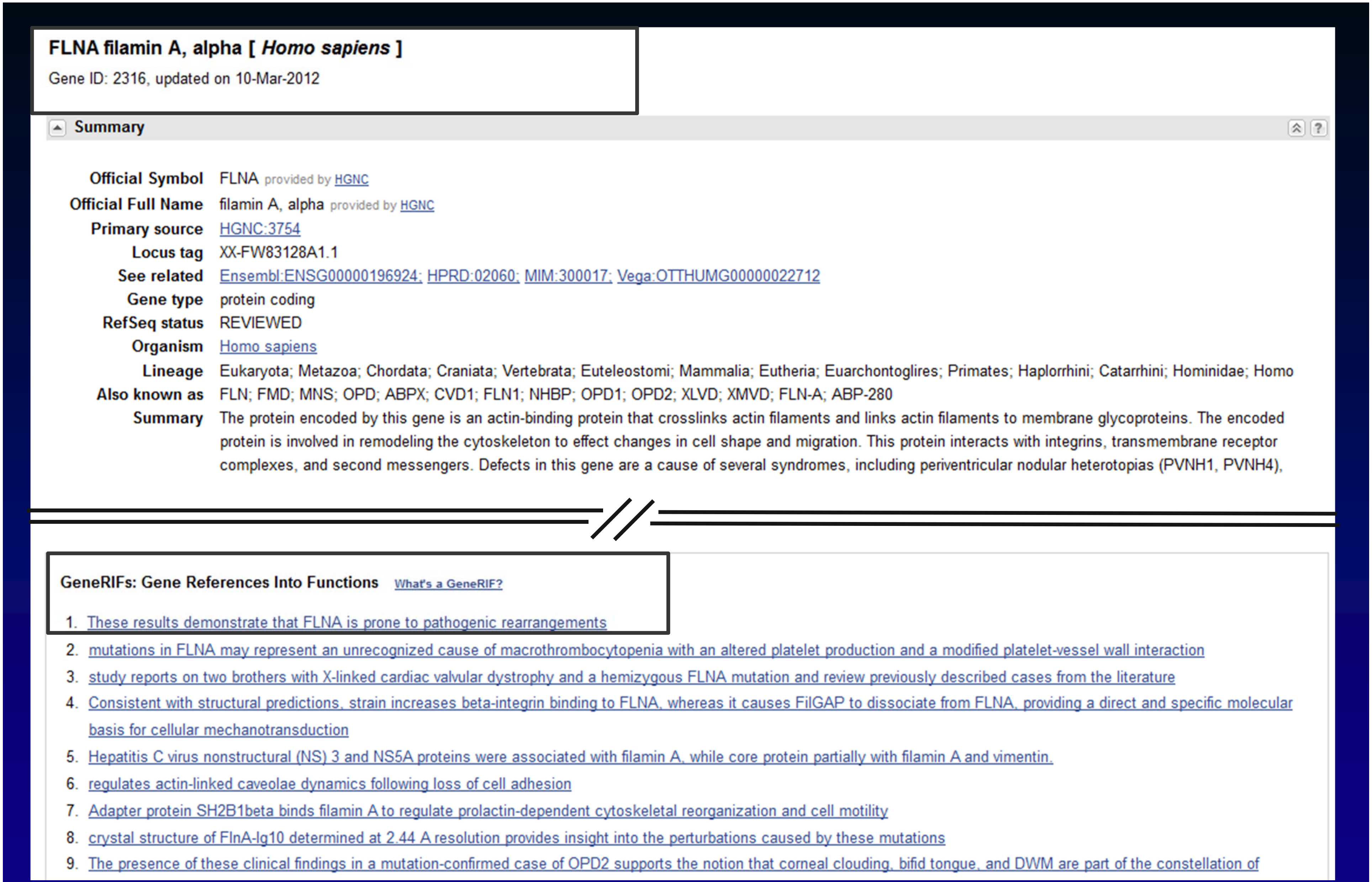
CheckTags Machine Learning Results

- 200k citations for training and 100k citations for testing

CheckTag	F_1 before ML	F_1 with ML	Improvement
Middle Aged	**1.01%**	**59.50%**	**+58.49**
Aged	11.72%	54.67%	+42.95
Child, Preschool	6.11%	45.40%	+39.29
Adult	19.49%	56.84%	+37.35
Male	38.47%	71.14%	+32.67
Aged, 80 and over	1.50%	30.89%	+29.39
Young Adult	2.83%	31.63%	+28.80
Female	46.06%	73.84%	+27.78
Adolescent	24.75%	42.36%	+17.61
Humans	79.98%	91.33%	+11.35
Infant	34.39%	44.69%	+10.30
Swine	71.04%	74.75%	+3.71

Research - J. Caitlin Sticco

- Introduction to Gene Indexing
- The Gene Indexing Assistant



FLNA filamin A, alpha [*Homo sapiens*]

Gene ID: 2316, updated on 10-Mar-2012

Summary

Official Symbol	FLNA provided by HGNC
Official Full Name	filamin A, alpha provided by HGNC
Primary source	HGNC:3754
Locus tag	XX-FW83128A1.1
See related	Ensembl:ENSG00000196924; HPRD:02060; MIM:300017; Vega:OTTHUMG00000022712
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	FLN; FMD; MNS; OPD; ABPX; CVD1; FLN1; NHBP; OPD1; OPD2; XLVD; XMVD; FLN-A; ABP-280
Summary	The protein encoded by this gene is an actin-binding protein that crosslinks actin filaments and links actin filaments to membrane glycoproteins. The encoded protein is involved in remodeling the cytoskeleton to effect changes in cell shape and migration. This protein interacts with integrins, transmembrane receptor complexes, and second messengers. Defects in this gene are a cause of several syndromes, including periventricular nodular heterotopias (PVNH1, PVNH4),

GeneRIFs: Gene References Into Functions What's a GeneRIF?

1. These results demonstrate that FLNA is prone to pathogenic rearrangements
2. mutations in FLNA may represent an unrecognized cause of macrothrombocytopenia with an altered platelet production and a modified platelet-vessel wall interaction
3. study reports on two brothers with X-linked cardiac valvular dystrophy and a hemizygous FLNA mutation and review previously described cases from the literature
4. Consistent with structural predictions, strain increases beta-integrin binding to FLNA, whereas it causes FilGAP to dissociate from FLNA, providing a direct and specific molecular basis for cellular mechanotransduction
5. Hepatitis C virus nonstructural (NS) 3 and NS5A proteins were associated with filamin A, while core protein partially with filamin A and vimentin.
6. regulates actin-linked caveolae dynamics following loss of cell adhesion
7. Adapter protein SH2B1beta binds filamin A to regulate prolactin-dependent cytoskeletal reorganization and cell motility
8. crystal structure of FlnA-Ig10 determined at 2.44 A resolution provides insight into the perturbations caused by these mutations
9. The presence of these clinical findings in a mutation-confirmed case of OPD2 supports the notion that corneal clouding, bifid tongue, and DWM are part of the constellation of

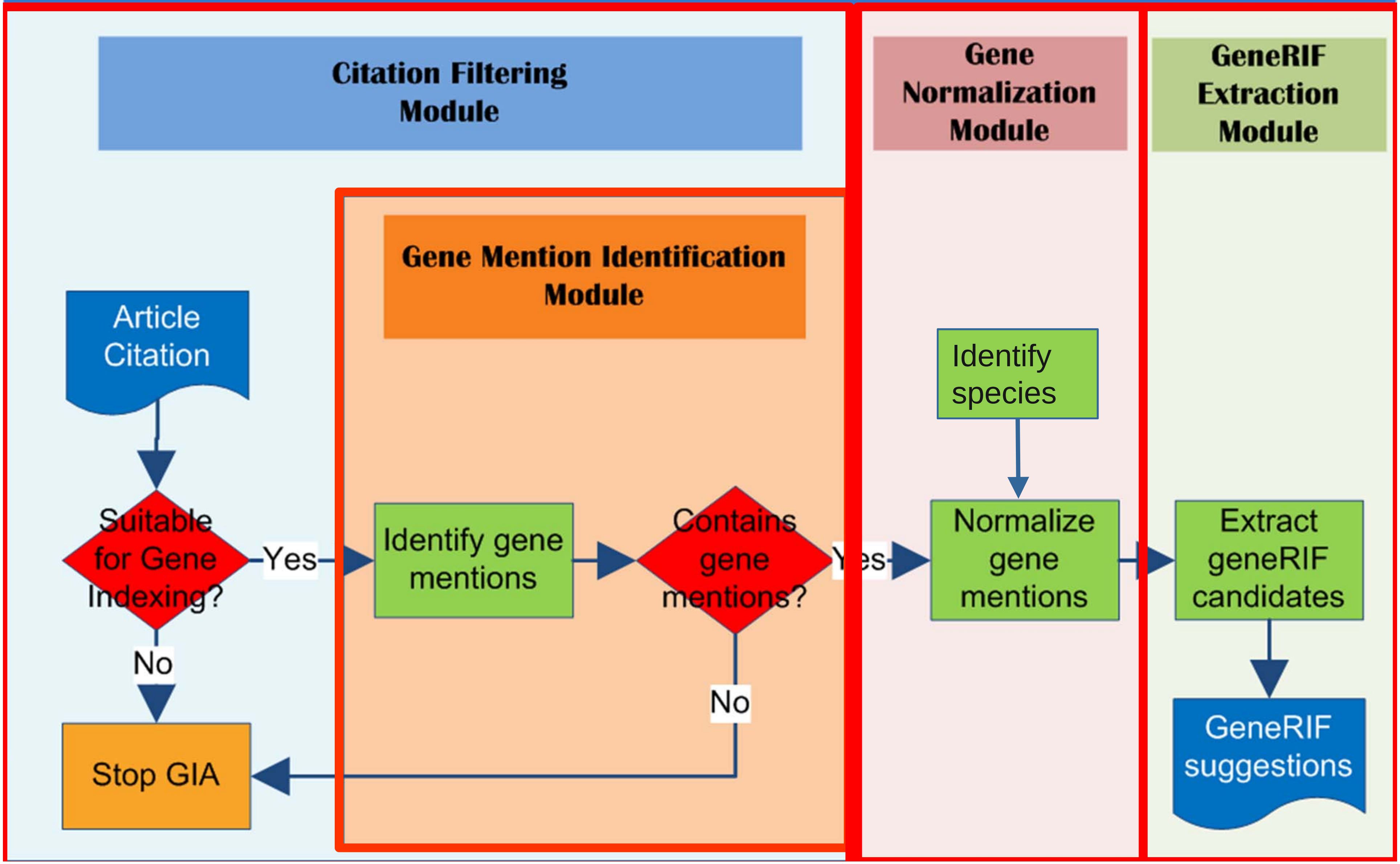
The Gene Indexing Assistant

- An automated tool to assist the indexer in identifying and creating GeneRIFs
  - Evaluate the article
  - Identify genes
  - Make links to Entrez Gene
  - Suggest geneRIF annotation
- Anticipated Benefits:
  - Increase in speed
  - Increase in comprehensiveness

Corpus Creation

- Gene mentions
  - tagged by manually correcting the automated program
- GeneRIF classes
  - Non-geneRIF, Structure, Function, Expression, Isolation, Reference, and Other
- Claims classes
  - Putative, Established, or Non-claim
- Discourse classes
  - Title, Background, Purpose, Methods, Results, Conclusions
  - Alternate dataset of 600,000 structured abstracts with similar labels

Gene Indexing Assistant Structure


Citation Filtering Module
Gene Normalization Module
GeneRIF Extraction Module
Gene Mention Identification Module
Article Citation
Suitable for Gene Indexing?
Yes
No
Stop GIA
Identify gene mentions
Contains gene mentions?
Yes
No
Identify species
Normalize gene mentions
Extract geneRIF candidates
GeneRIF suggestions

Software Origins

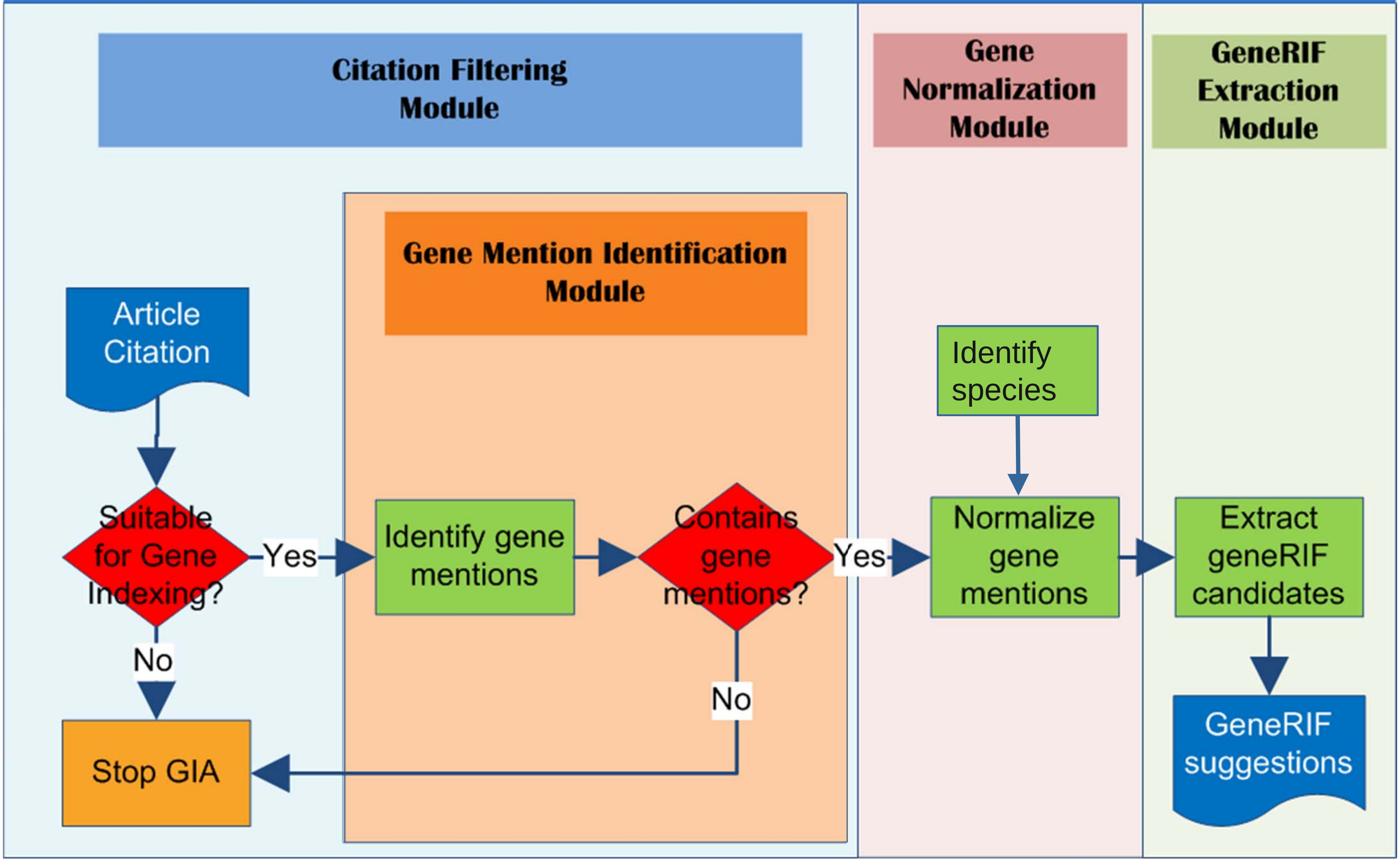
Integrated External Software

- GNAT from Jorg Hakenberg
  - Include BANNER for gene identification
- Linnaeus from Gerner, Nenadic, and Bergman
- Organism Tagger from Naderi et al.

Components Developed In-house

- Framework
- Hand-curated dictionary
- In-house modules for human gene identification, normalization, and geneRIF extraction

Gene Indexing Assistant Structure


Citation Filtering Module
Gene Normalization Module
GeneRIF Extraction Module
Gene Mention Identification Module
Article Citation
Suitable for Gene Indexing?
Yes
No
Identify gene mentions
Contains gene mentions?
Yes
No
Stop GIA
Identify species
Normalize gene mentions
Extract geneRIF candidates
GeneRIF suggestions

Gene Mention Identification

Filamin a mediates HGF/c-MET signaling in tumor cell migration.

Deregulated hepatocyte growth factor (HGF)/c-MET axis has been correlated with poor clinical outcome and drug resistance in many human cancers. In our study, we show that multiple human cancer tissues and cells express filamin A (FLNA), a large cytoskeletal actin-binding protein, and expression of c-MET is significantly reduced in human tumor cells deficient for FLNA.

Gene Mention Identification

Filamin a mediates HGF/c-MET signaling in tumor cell migration.

Deregulated hepatocyte growth factor (HGF)/c-MET axis has been correlated with poor clinical outcome and drug resistance in many human cancers. In our study, we show that multiple human cancer tissues and cells express filamin A (FLNA), a large cytoskeletal actin-binding protein, and expression of c-MET is significantly reduced in human tumor cells deficient for FLNA.

filamin a, flna, hepatocycte growth factor, c-met

Gene Mention Identification

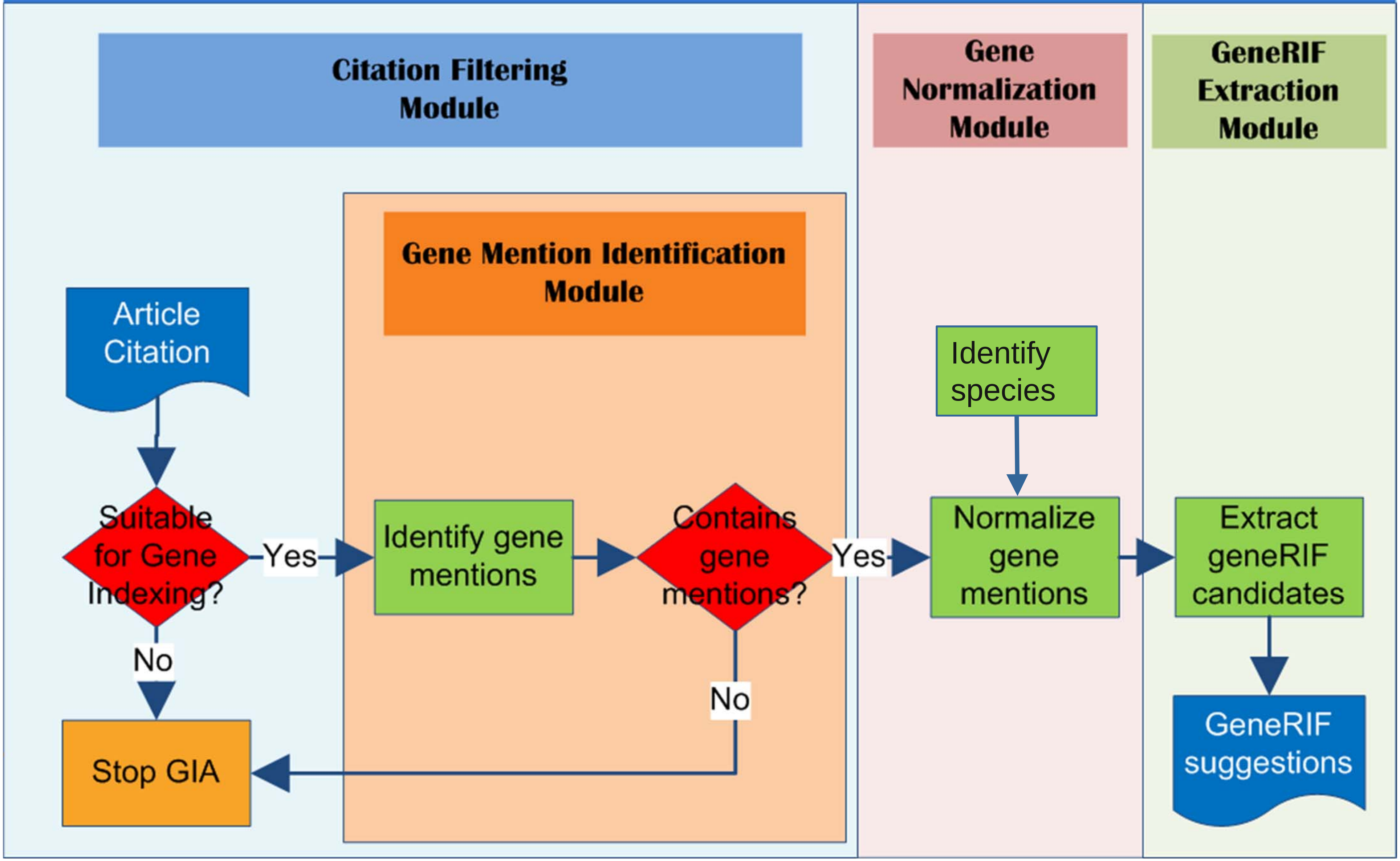
In-House Components

- Hand curated dictionary
  - Derived from Entrez Gene
  - Filtering for problem synonyms
  - Variant creation (reductive tokenization?)
- Strict Dictionary Mapping

External Components

- GNAT: Conditional Random Fields (CRF) from BANNER

Gene Indexing Assistant Structure



Citation Filtering Module

Gene Normalization Module

GeneRIF Extraction Module

Gene Mention Identification Module

Article Citation

Suitable for Gene Indexing?

Yes

No

Stop GIA

Identify gene mentions

Contains gene mentions?

Yes

No

Identify species

Normalize gene mentions

Extract geneRIF candidates

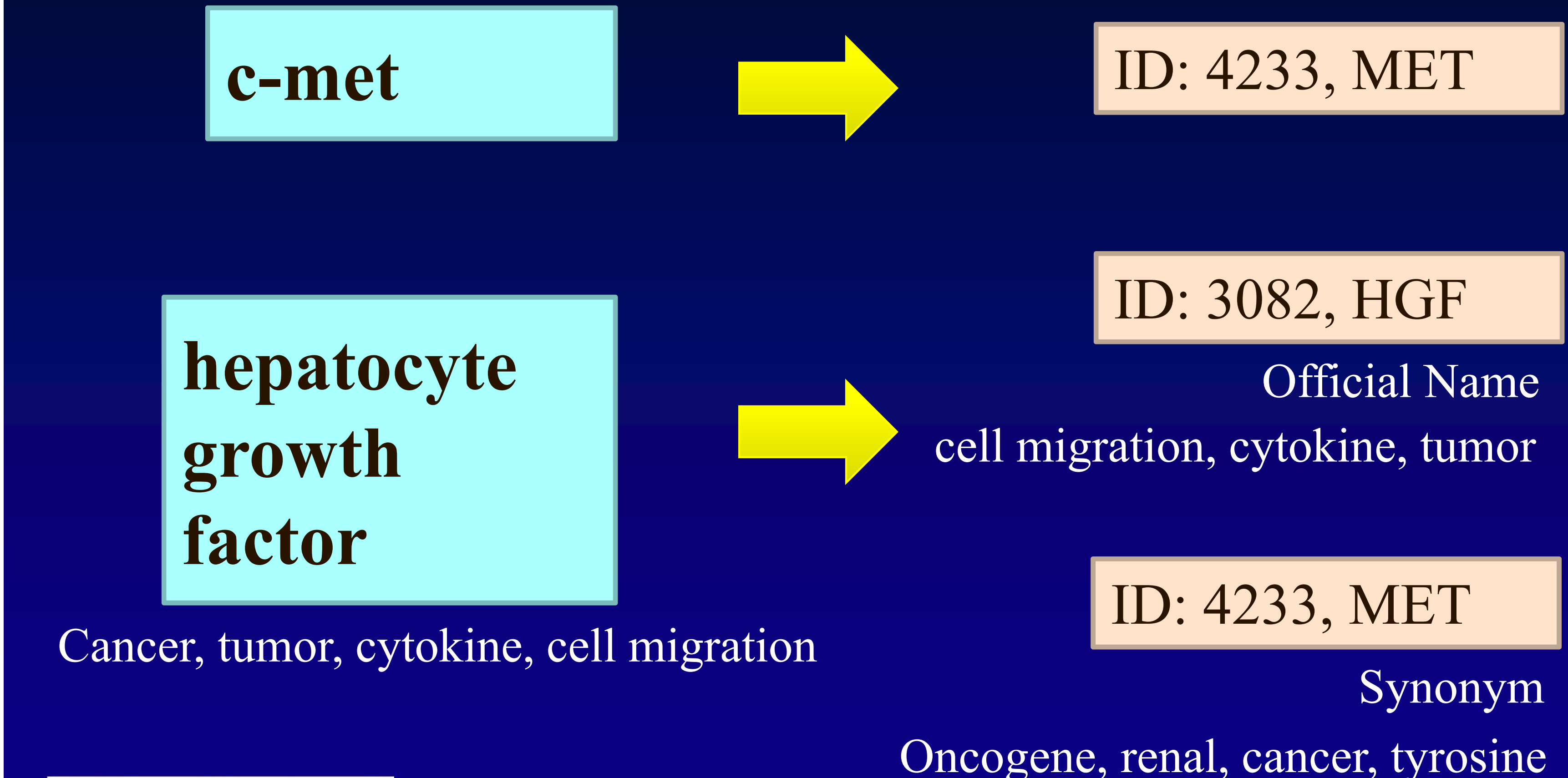
GeneRIF suggestions

Species Identification and Assignment

External Components

- Identification
  - Linnaeus: includes common names and maps stand alone genera to most likely species
  - Organism Tagger: includes cell lines and microbial strains
- Assigning genes to species
  - GNAT: Proximity heuristic

Gene Mention Normalization



c-met

ID: 4233, MET

hepatocyte growth factor

Cancer, tumor, cytokine, cell migration

ID: 3082, HGF

Official Name

cell migration, cytokine, tumor

ID: 4233, MET

Synonym

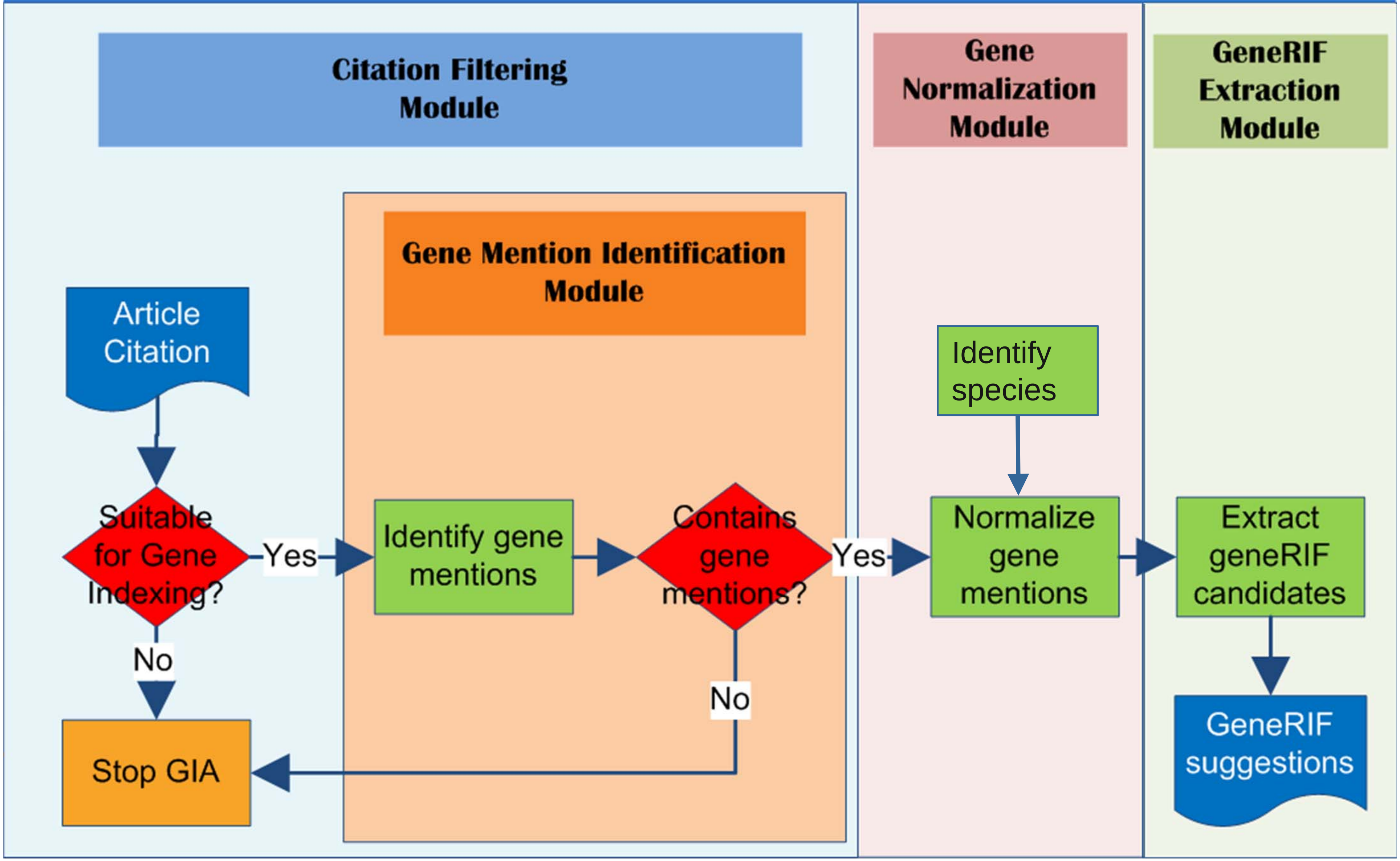
Oncogene, renal, cancer, tyrosine

Gene Mention Normalization

Identification and Normalization Results

Species	Recall	Precision	F_1
Human	**83%**	**80%**	**81%**

Gene Indexing Assistant Structure



Citation Filtering Module

Gene Normalization Module

GeneRIF Extraction Module

Gene Mention Identification Module

Article Citation

Suitable for Gene Indexing?

Yes

No

Stop GIA

Identify gene mentions

Contains gene mentions?

Yes

No

Identify species

Normalize gene mentions

Extract geneRIF candidates

GeneRIF suggestions

Classifier Results

Features	Precision	Recall	F_1
Position (pos)	72%	73%	72%
Text (word features)	63%	64%	63%
Gene Names	55%	70%	62%
Discourse (Structured Ab. Labels)	70%	80%	75%
pos + discourse	70%	86%	76.89%
★ pos + discourse + GO	70%	86%	77.07%

Future Improvements and Research Areas

- Additional preprocessing
  - Expand certain anaphora
- Extracting interaction data
- Expanding the dictionaries
- Improved abbreviation resolution
- Additional training for low-performing species
- Integration of additional identification or normalization software

Research and Outreach Efforts (concl.)

- External Collaboration
  - IBM DeepQA group: applying Watson to health care
- Data Dissemination
  - MEDLINE Baseline Repository
  - WSD test collections
- Biomedical NLP/IR Challenges
  - Text Retrieval Conference (TREC)
    - Genomics track
    - Medical Records track
  - Informatics for Integrating Biology & the Bedside (i2b2)
  - Medical NLP Challenge

Tomorrow …

LHNCBC Participation in NLP/IR Challenges

Outline

- Introduction [Lan]
- MetaMap [François]
- The NLM Medical Text Indexer (MTI) [Jim]
- Availability of Indexing Initiative Tools [Willie]
- Research and Outreach Efforts [Antonio, Caitlin, Lan]
- Summary and Future Plans [Lan]
- Questions

Indexing Initiative Top 10 (1/2)

10. ‘MTI Why’ explanation facility

9. Application of MTI to Cataloging and History of Medicine records

8. The MetaMap UIMA wrapper, increasing MetaMap’s availability

7. Significant speedup of MetaMap

6. Collaboration with IBM DeepQA group applying Watson to health care

Indexing Initiative Top 10 (2/2)

5. The development of Gene Indexing Assistant (GIA)

4. More WSD methods with better results

3. Improvement in MTI’s performance due to technical enhancements and close collaboration with Index Section

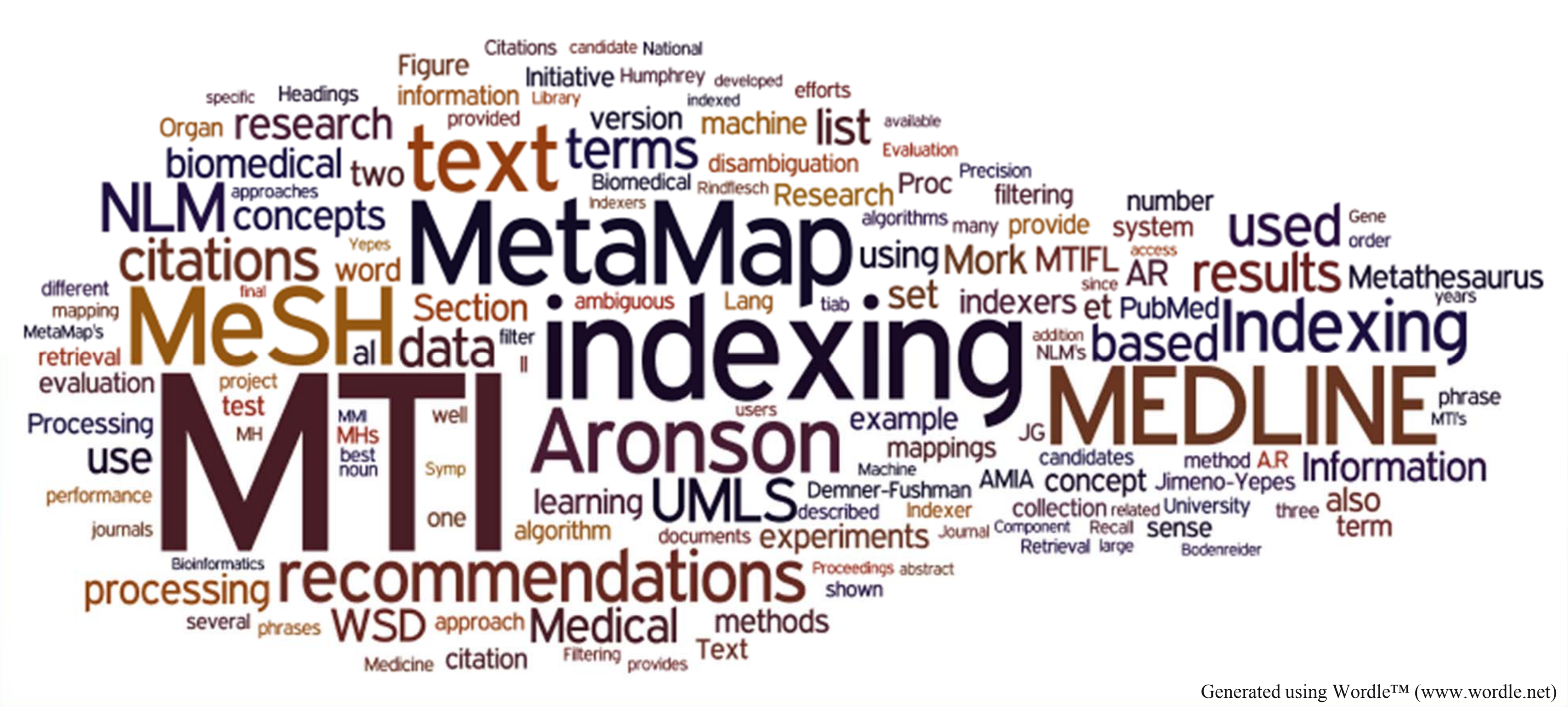
2. Downloadable releases of MetaMap, especially for Windows

Inauguration of MTI as a first-line indexer (MTIFL)!

Future Plans

- Continued collaboration with
  - The NLM Index Section
  - IBM and other external organizations
- Planned improvements to MetaMap and MTI such as
  - Expansion/improvement of MTIFL capability
  - Add species detection to MTI for disambiguation and for GIA
  - Further MTI research with Antonio Jimeno-Yepes and Caitlin Sticco
  - Possible high-level MetaMap modularization to facilitate plug and play strategies

Questions


Citations candidate National
Figure Initiative Humphrey developed efforts
specific Headings information Library indexed
Organ research provided version machine list available
biomedical two text terms disambiguation Evaluation Precision
NLM approaches Biomedical Rindflesch Research Proc filtering
concepts Indexers algorithms number
Yepes MetaMap many provide system used Gene order
citations word using Mork MTIFL access AR results Metathesaurus
different final since years
mapping Section ambiguous Lang tiab set indexers et PubMed
MetaMap's MeSH addition
retrieval all data filter indexing NLM's based Indexing
II
evaluation project MEDLINE phrase
test MTI users MTI's
Processing MMI well example
MHs MH JG
best mappings
noun Aronson candidates
use Machine method AR Information
Symp AMIA concept Jimeno-Yepes
performance learning UMLS Demner-Fushman
described Indexer collection related University three also
one
journals algorithm documents experiments Journal Component Recall sense term
Retrieval large Bodenreider
Bioinformatics Proceedings abstract
processing recommendations shown
several phrases WSD approach Medical methods
Medicine citation Filtering provides Text

Generated using Wordle™ (www.wordle.net)

Alan (Lan) R. Aronson
James G. Mork
François-Michel Lang

Willie J. Rogers
Antonio J. Jimeno-Yepes
J. Caitlin Sticco

Extra slides in case of questions

Candidate Pruning: Output Example

protein-4 FN3 fibronectin type III domain GSH glutathione GST glutathione S-transferase hIL-6 human interleukin-6 HSA human serum albumin IC(50) half-maximal inhibitory concentration Ig immunoglobulin IMAC immobilized metal affinity chromatography K(D) equilibrium constant

Candidate Pruning: Output Example

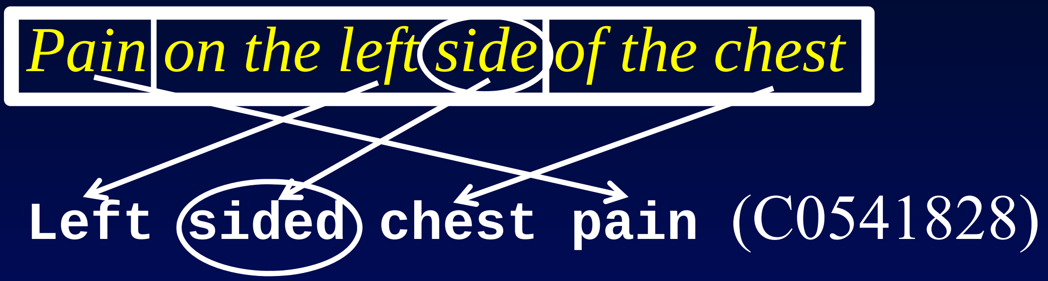
```
(Total=99; Excluded=13; Pruned=50; Remaining=36)
783   equilibrium constant [npop]
780 P Equilibrium [orgf]
780 P Kind of quantity - Equilibrium [qnco]
780 P Constant (qualifier) [qlco]
713   protein K [aapp]
691   Protein concentration [lbpr]
671   protein serum [aapp,bacs]
671   Protein.serum [lbtr]
656 P serum K+ [lbpr]
656   protein human [aapp,bacs]
653   Human immunoglobulin [aapp,imft,phsu]
```

User-Defined Acronyms (UDAs)

Simply create a text file with UDA definitions:

```
CABG | coronary artery bypass graft
PTCA | percutaneous transluminal coronary angioplasty
RBBB | right bundle branch block
LAFB | left anterior fascicular block
AV   | aortic valve
PTLD | post-transplantation lymphoproliferative disorder
CHOP | cyclophosphamide, hydroxydaunomycin, Oncovin, and prednisone
LIMA | left internal mammary artery
LAD  | left anterior descending coronary artery
SVG  | saphenous vein graft
PLB  | posterolateral bundle
PDA  | posterior descending artery
IM   | internal mammary
```

Complexity - Composite Phrases



Pain on the left side of the chest

Left sided chest pain (C0541828)

Linguistic variants

Syntactic processing

Word order

10^{21} Terabytes of Memory?!

$10^{21} = 10^{10} * 10^{11}$

$= \underbrace{(10 \text{ billion})}_{\text{150\% of world population}} * \underbrace{(100 \text{ billion})}_{\text{Required terabytes/person}}$

Oak Ridge National Lab’s Cray Jaguar: 300TB

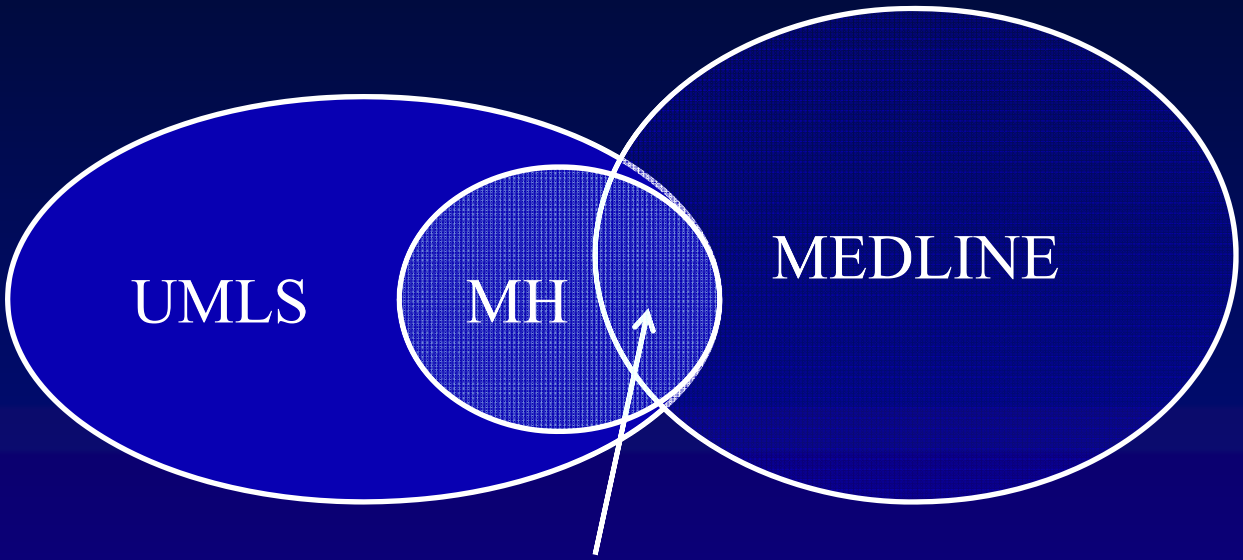
Concepts with at least 300 Synonyms

- 349: C1163679|Water 1000 MG/ML Injectable Solution

- 327: C0874083|Triclosan 3 MG/ML Medicated Liquid Soap

- 312: C0980221|Sodium Chloride 0.154 MEQ/ML Injectable Solution

MSH WSD corpus



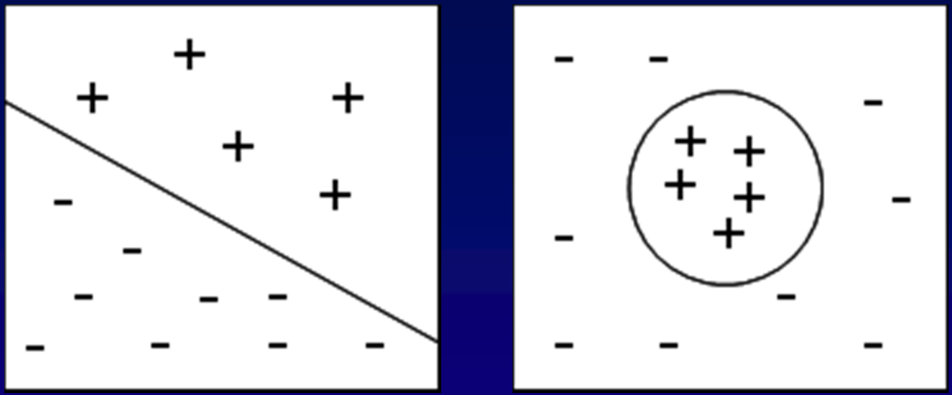
UMLS

MH

MEDLINE

Disambiguation corpus

Meta-learning



ML: Human MeSH heading

Method	Average F-measure
MTI	0.72
Naïve Bayes	0.85
Support vector machine	0.88
AdaBoostM1	0.92

Accuracy

- Accuracy is how close a measured value is to the **actual (true) value**

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- Precision, proportion of relevant predictions

$$precision = \frac{TP}{TP + FP}$$

Micro/macro averaging

- Macro averaging takes into account the category (MH)
- Micro averaging does not consider MH

MH	True Pos	False Pos	Positive	Precision	Recall	F-measure
Humans	66,429	5,985	71,484	0.9174	0.9293	0.9233
Male	24,664	7,107	34,463	0.7763	0.7157	0.7448
Female	25,824	6,718	35,501	0.7936	0.7274	0.7590
Macro				0.8291	0.7908	0.8090
Micro	116,917	19,810	141,448	0.8551	0.8266	0.8406

MetaMap Indexing (MMI)

- Summarizes and scores what is found within a citation
- Location - Title given more emphasis
- Frequency of occurrence
- Relevancy:
  - MeSH Tree Depth
  - MetaMap score
- Provides a scored and ordered list of UMLS concepts describing the citation
- Provides our best indicator of MeSH Headings

Restrict to MeSH

- Allows us to map UMLS concepts to MeSH Headings
- Maps nomenclature to MeSH

Encephalitis Virus, California

ET: Jamestown Canyon virus

ET: Tahyna virus

Inkoo virus

Jerry Slough virus

Keystone virus

Melao virus

San Angelo virus

Serra do Navio virus

Snowshoe hare virus

Trivittatus virus

Lumbo virus

South River virus

California Group Viruses

PubMed Related Citations (PRC)

- Uses PubMed pre-calculated related articles, same as DCMS Related Articles tab
- Provides terms not available in title/abstract
- Used to filter and support MeSH Headings identified by MetaMap Indexing
- Only use MeSH Headings, no CheckTags, no Subheadings, no Supplemental Concepts
- Can provide non-related terms, so heavily filtered

MTI – Initial MTIFL Journals (Feb 18, 2011)

BMC Bioinformatics
Journal of Industrial Microbiology & Biotechnology
Letters in Applied Microbiology
GPB Genomics, Proteomics & Bioinformatics
Archives of Microbiology
environmental microbiology
Journal of morphology
Journal of Applied Microbiology
FEMS Microbiology Ecology
Bioinformatics
International Journal of Food Microbiology
veterinary microbiology
The ISME Journal
Multidisciplinary Journal of Microbial Ecology
Canadian Journal of Microbiology
Revue canadienne de microbiologie

MTI – Added MTIFL Journals

Added June 1, 2011

ENVIRONMENT INTERNATIONAL
A Journal of Environmental Science, Risk & Health

(17)

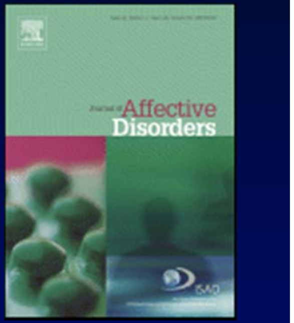
Added August 18, 2011

Added September 5, 2011

(19)

MTI – Added MTIFL Journals

Added October 5, 2011



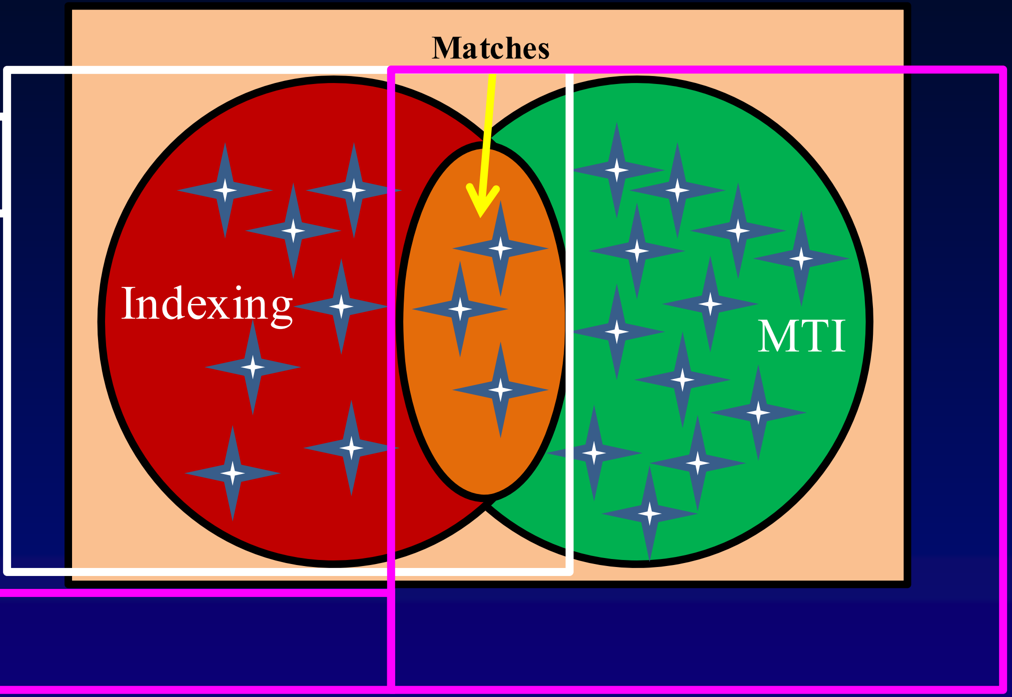
(23)

MTIFL Journal Performance

	Current MTIFL Statistics						Previous Results							
	2012				Diff 2011	Diff 2010	2010				2011			
Journal	Articles	Recall	Precision	F_1			Articles	Recall	Precision	F_1	Articles	Recall	Precision	F_1
Arch Microbiol	15	57.24%	58.78%	58.00%	-1.15%	4.67%	103	71.50%	44.01%	54.48%	69	55.97%	62.71%	59.15%
Bioinformatics	113	57.98%	64.65%	61.13%	2.38%	15.75%	820	76.61%	29.89%	43.01%	433	53.66%	64.91%	58.75%
BMC Bioinformatics	126	63.84%	70.80%	67.14%	5.33%	19.69%	851	77.83%	28.87%	42.12%	403	57.13%	67.33%	61.81%
Can J Microbiol	22	60.53%	61.69%	61.10%	-1.37%	16.22%	131	67.07%	35.29%	46.25%	59	61.07%	63.94%	62.47%
Curr Opin Biotechnol	29	73.81%	79.15%	76.39%	7.99%	38.45%	99	53.86%	20.73%	29.94%	25	59.73%	80.00%	68.39%
Curr Opin Cell Biol	0	0.00%	0.00%	0.00%	0.00%	33.39%	97	54.38%	26.60%	35.72%	31	70.94%	67.38%	69.12%
Ecotoxicol Environ Saf	42	69.91%	79.74%	74.50%	2.81%	27.96%	122	68.92%	32.03%	43.73%	199	65.42%	79.30%	71.69%
Environ Int	11	68.21%	77.44%	72.54%	7.47%	22.33%	92	55.94%	34.57%	42.73%	54	57.20%	75.44%	65.06%
Environ Microbiol	58	60.92%	71.62%	65.84%	3.55%	13.98%	256	63.68%	38.91%	48.31%	183	58.54%	66.56%	62.29%
Environ Toxicol	15	75.26%	76.88%	76.06%	5.98%	25.44%	49	68.25%	33.17%	44.65%	24	63.73%	77.85%	70.08%
Environ Toxicol Chem	54	68.00%	72.27%	70.07%	1.87%	22.42%	287	66.24%	34.98%	45.78%	111	62.44%	75.13%	68.20%
FEMS Microbiol Ecol	0	0.00%	0.00%	0.00%	0.00%	8.60%	178	68.62%	44.11%	53.70%	157	58.32%	66.87%	62.30%
Genomics Proteomics Bioinformatics	0	0.00%	0.00%	0.00%	0.00%	7.29%	30	77.30%	35.80%	48.93%	15	50.36%	63.64%	56.22%
Health Psychol	20	80.36%	74.75%	77.45%	8.06%	30.29%	93	45.06%	34.54%	39.11%	18	67.08%	71.88%	69.40%
Int J Food Microbiol	12	81.89%	74.82%	78.20%	14.82%	14.57%	305	69.95%	37.48%	48.81%	272	62.48%	64.31%	63.38%
ISME J	34	64.02%	62.69%	63.35%	1.80%	15.78%	122	65.03%	35.31%	45.77%	120	58.00%	65.56%	61.55%
J Affect Disord	130	82.60%	91.44%	86.80%	50.47%	New	338	45.32%	30.32%	36.33%	0	0.00%	0.00%	0.00%
J Appl Microbiol	49	59.33%	65.36%	62.20%	-0.16%	16.19%	562	71.73%	34.04%	46.17%	489	60.38%	64.48%	62.36%
J Ind Microbiol Biotechnol	26	71.21%	81.31%	75.93%	10.35%	19.66%	107	66.90%	34.95%	45.92%	82	64.23%	66.98%	65.58%
J Morphol	30	76.34%	62.31%	68.61%	-1.22%	28.91%	131	65.02%	29.85%	40.92%	64	76.85%	63.98%	69.83%
Lett Appl Microbiol	60	64.14%	69.27%	66.61%	-0.06%	15.00%	188	71.46%	40.46%	51.67%	116	65.13%	68.28%	66.67%
Nord J Psychiatry	19	79.32%	72.76%	75.90%	-4.61%	42.83%	55	43.37%	33.30%	37.68%	9	79.17%	81.90%	80.51%
Vet Microbiol	25	79.69%	72.73%	76.05%	10.01%	18.82%	285	71.54%	35.24%	47.22%	278	64.54%	67.61%	66.04%
Totals	890	69.99%	74.85%	72.34%	8.35%	19.67%	5,301	66.64%	33.19%	44.31%	3,211	60.74%	67.60%	63.99%

Precision, Recall, F-Measure

Recall: 3/10 = 0.3



10 Indexing

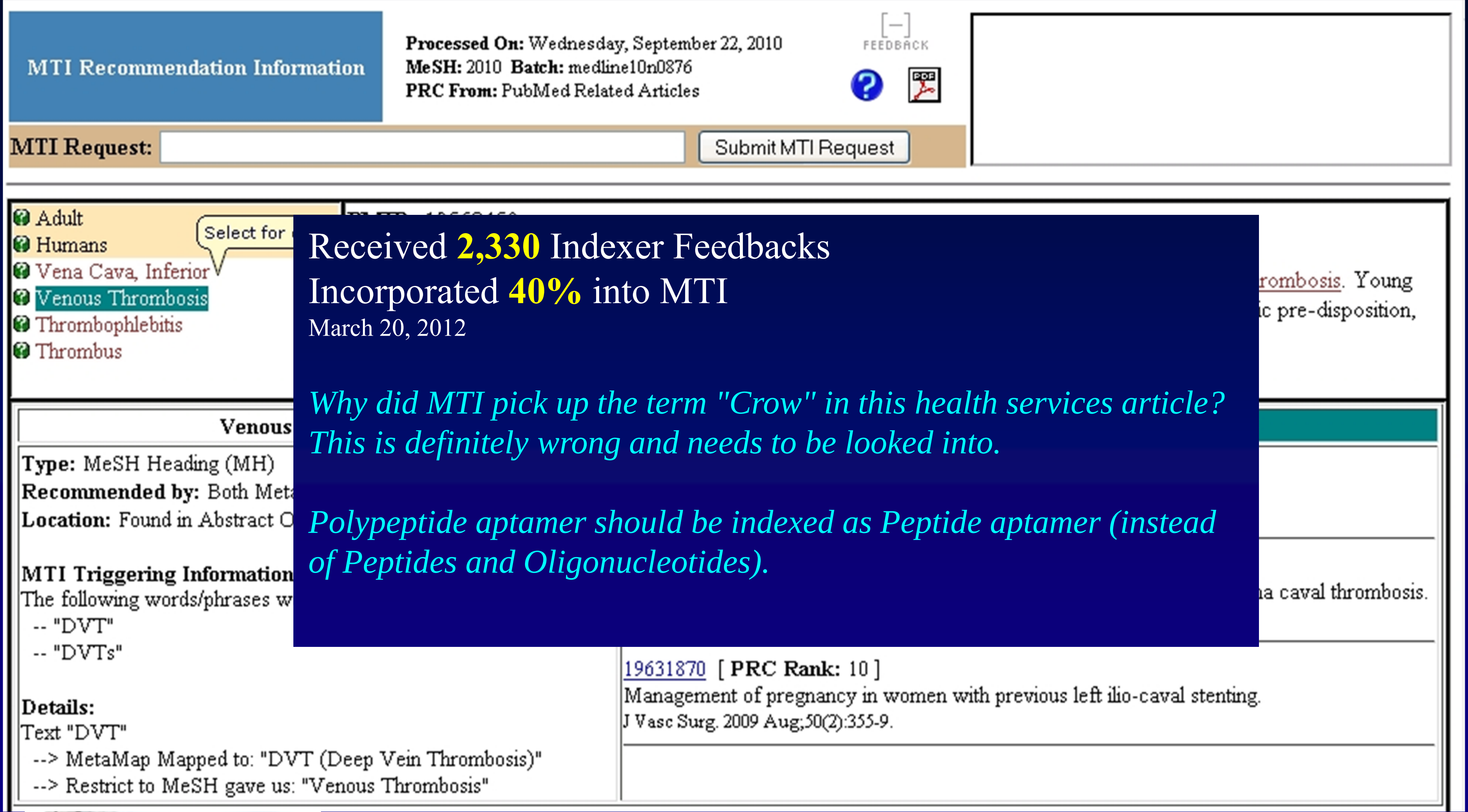
15 MTI

3 Matches

Precision: 3/15 = 0.2

F_1-Measure: (2 * 0.2 * 0.3) / (0.2 + 0.3) = 0.24

MTIWhy



MTI Recommendation Information

Processed On: Wednesday, September 22, 2010
MeSH: 2010 **Batch:** medline10n0876
PRC From: PubMed Related Articles

MTI Request:

- Adult
- Humans
- Vena Cava, Inferior
- Venous Thrombosis
- Thrombophlebitis
- Thrombus

rombosis. Young
ic pre-disposition,

Venous

Type: MeSH Heading (MH)
Recommended by: Both Meta
Location: Found in Abstract O

MTI Triggering Information
The following words/phrases w
-- "DVT"
-- "DVTs"

Details:
Text "DVT"
--> MetaMap Mapped to: "DVT (Deep Vein Thrombosis)"
--> Restrict to MeSH gave us: "Venous Thrombosis"

a caval thrombosis.

19631870 [**PRC Rank:** 10]
Management of pregnancy in women with previous left ilio-caval stenting.
J Vasc Surg. 2009 Aug;50(2):355-9.

Received **2,330** Indexer Feedbacks
Incorporated **40%** into MTI
March 20, 2012

Why did MTI pick up the term "Crow" in this health services article? This is definitely wrong and needs to be looked into.

Polypeptide aptamer should be indexed as Peptide aptamer (instead of Peptides and Oligonucleotides).

Questions

Alan (Lan) R. Aronson

James G. Mork

François-Michel Lang

Willie J. Rogers

Antonio J. Jimeno-Yepes

J. Caitlin Sticco