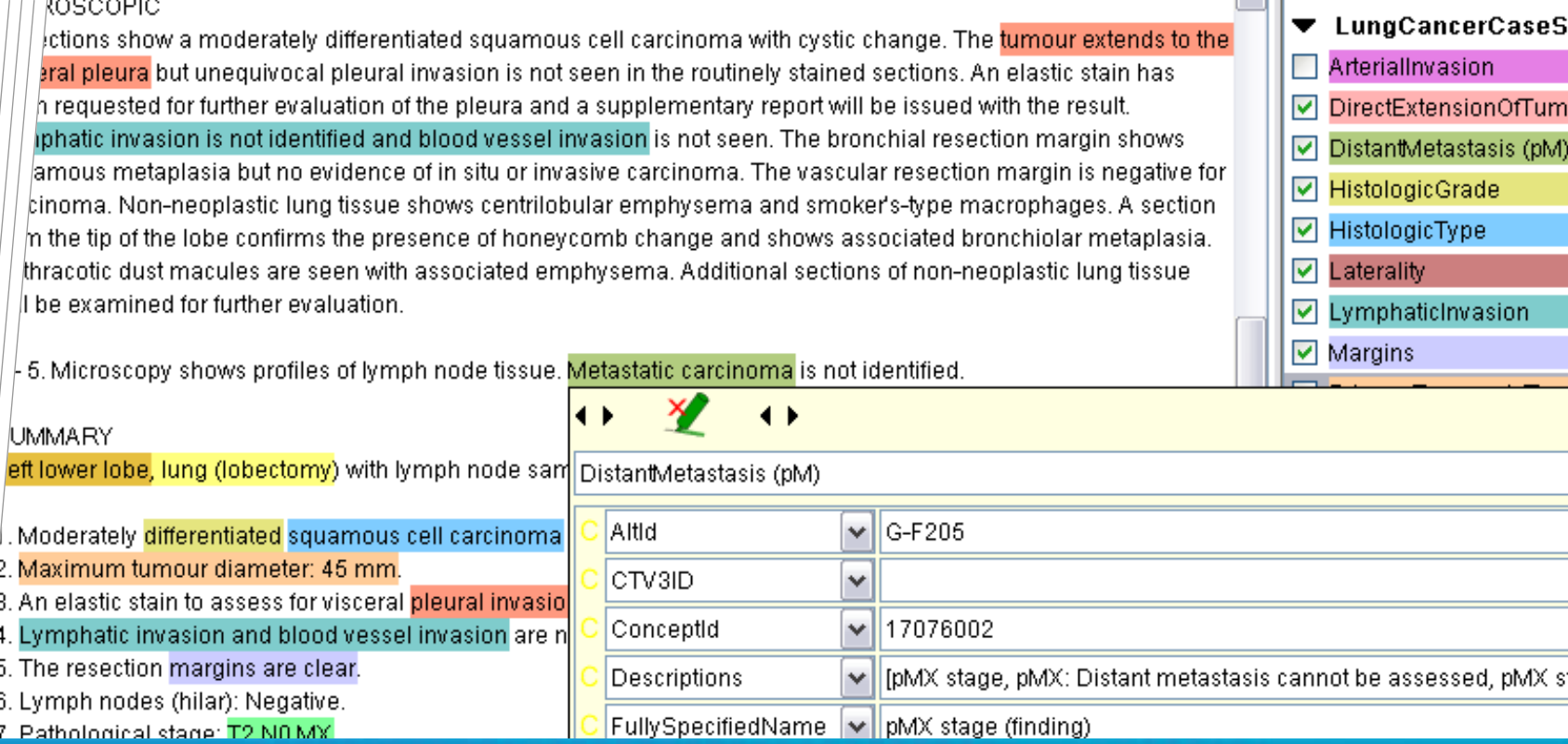


www.csiro.au

A Simple Pipeline Application for

Identifying and Negating SNOMED CT in Free Text

Anthony Nguyen[1], Michael Lawley[1], David Hansen[1], Shoni Colquist[2]

[1]The Australian e-Health Research Centre, CSIRO ICT Centre, Brisbane, Australia
[2]Queensland Cancer Control Analysis Team, Queensland Health, Brisbane, Australia

CSIRO

Medical Free Text to SNOMED CT Mapping

Medical free text

- Non-standardised
- Individual & institutional variation
- Manual extraction of information

SNOMED CT

- Standard reference terminology[1]
- Large formal ontology
- Aggregate clinical information for retrieval & analysis

[1] SNOMED CT is identified by NeHTA as the standard ontology to be used in systems within Australian Healthcare

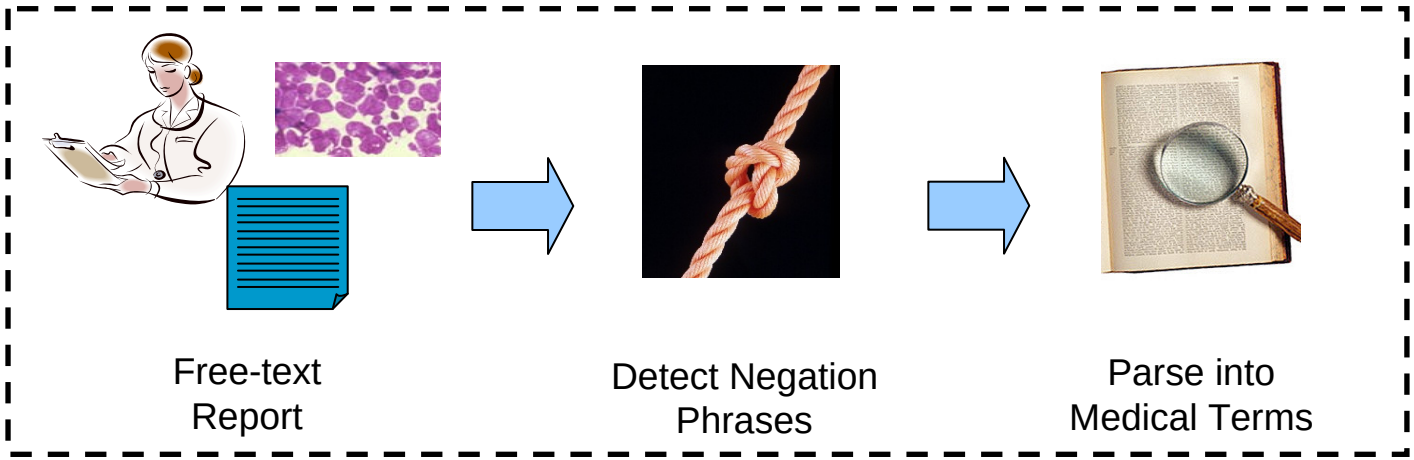
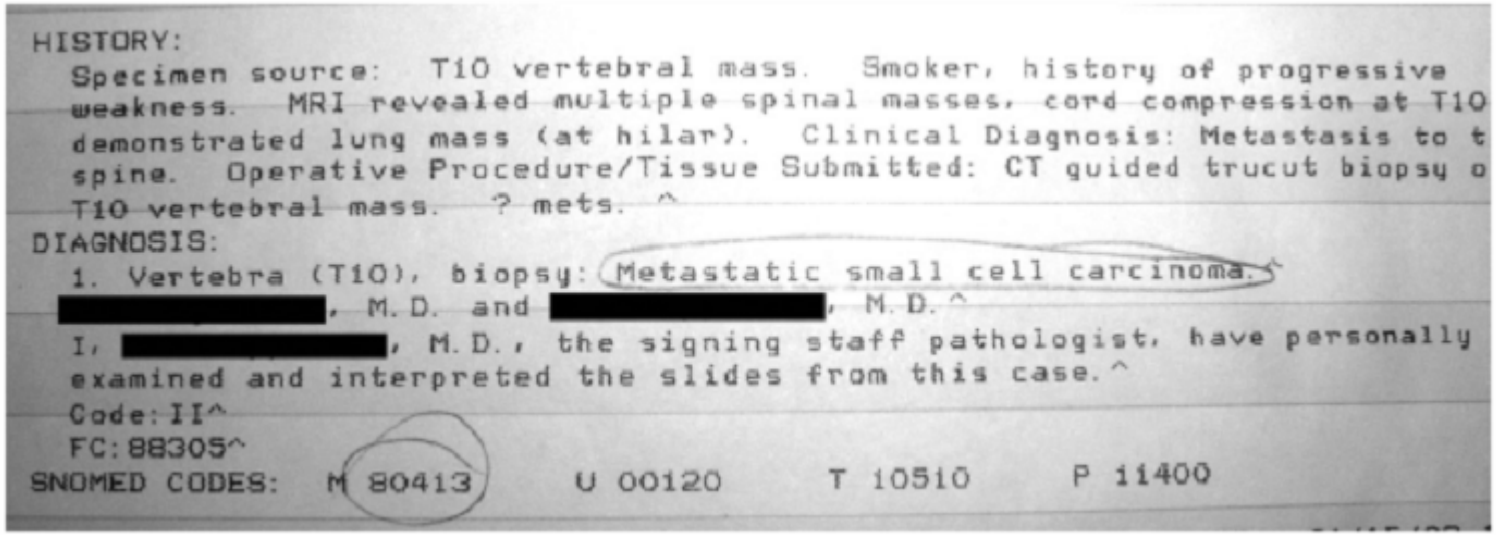


Figure 1. General methodology for coded biomedical terminology mapping from medical free text

Case Study: Cancer Notifications

- Manual review of (text based) medical records



HISTORY:
Specimen source: T10 vertebral mass. Smoker, history of progressive weakness. MRI revealed multiple spinal masses, cord compression at T10 demonstrated lung mass (at hilar). Clinical Diagnosis: Metastasis to t spine. Operative Procedure/Tissue Submitted: CT guided trucut biopsy o T10 vertebral mass. ? mets. ^
DIAGNOSIS:
1. Vertebra (T10), biopsy: Metastatic small cell carcinoma.
[redacted], M.D. and [redacted], M.D. ^
I, [redacted], M.D., the signing staff pathologist, have personally examined and interpreted the slides from this case. ^
Code: II^
FC: 88305^
SNOMED CODES: M 80413 U 00120 T 10510 P 11400

Figure 2. De-identified pathology report demonstrating how cancer notifications was historically performed

- Multi-year cancer information reporting delay
- Expensive and labour intensive process
- Subject to omission errors
  - Cases inadvertently skipped / keywords missed

CSIRO

Goal

- Hypothesis: Automated computer system could automatically perform the time and labour intensive manual coding of clinical information
- Method: Automatically scan free-text medical documents for clinically relevant terms

Pipeline Application [Tools]

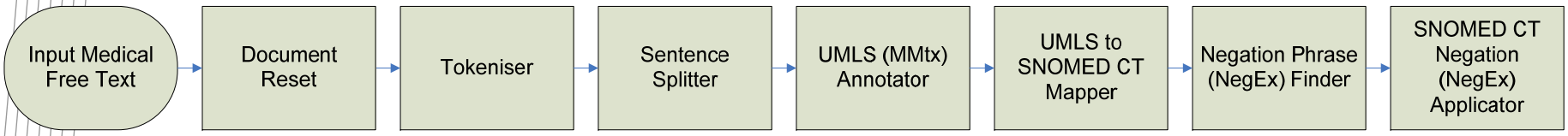
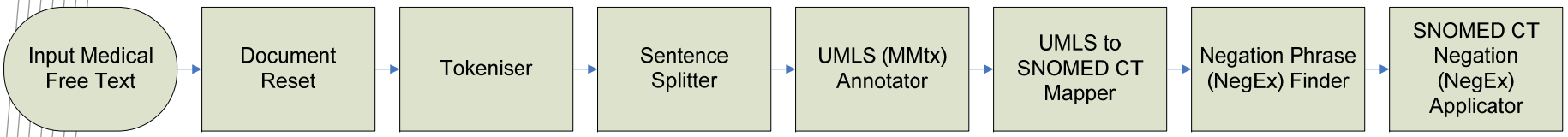


Figure 3. Pipeline application for the annotation of SNOMED CT concepts from free text

- General Architecture for Text Engineering (GATE)
  - Open source architecture for natural language processing (NLP)
- MetaMap Transfer (MMTx) application
  - Identify concepts in free text which are restricted to the SNOMED CT source
- Negation detection algorithm (NegEx)
  - Detection of negation phrases
- Ontology server (OntoServer)
  - SNOMED CT semantics to identify subsets of “findings” and “diseases” concepts for negation
- Development corpus (AUSLAB)
  - Lung cancer reports from state-wide pathology information system

Pipeline Application



- *Document Reset:*
  - Restores the document to its original state by deleting existing annotations

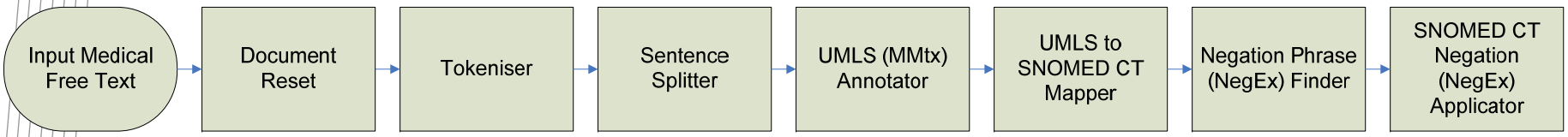
- *Tokeniser*:
  - Splits the document into:
    - Tokens
    - Length measurements & units
    - TNM cancer stages
    - Legacy SNOMED IDs
    - De-identified information

SUMMARY
Wedge resection of lesion in left lower lobe:
Mucinous bronchioloalveolar carcinoma, 12 mm in maximum diameter.
Pathological stage T1 Nx.

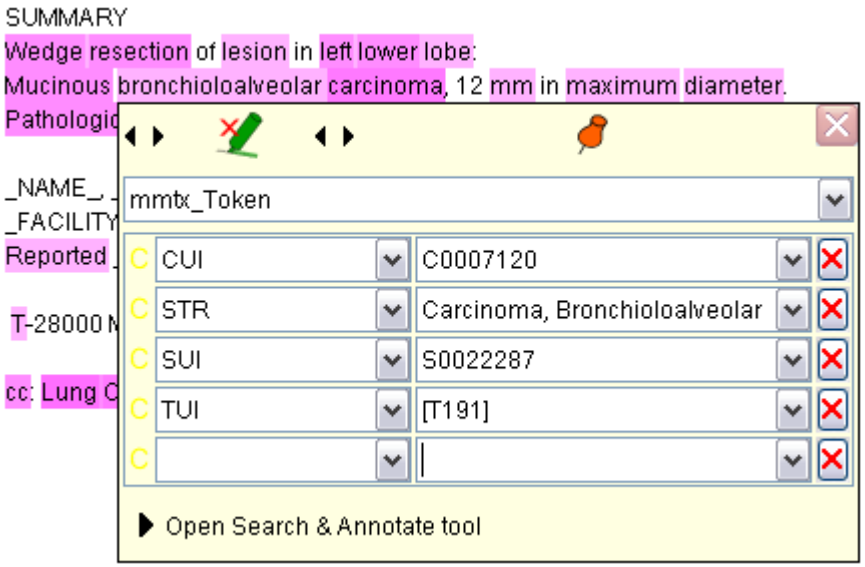
NAME, _NAME_
FACILITYORPLACE
Reported _DATE_

T-28000 M-82503 P1-03000

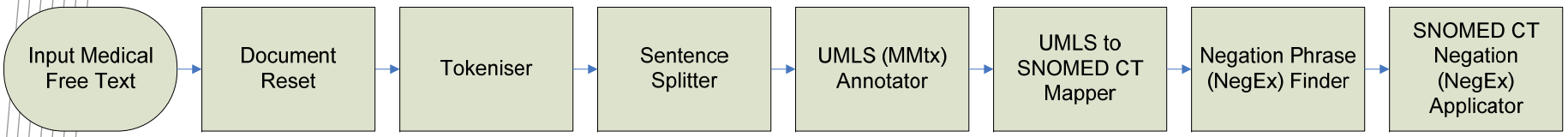
Pipeline Application



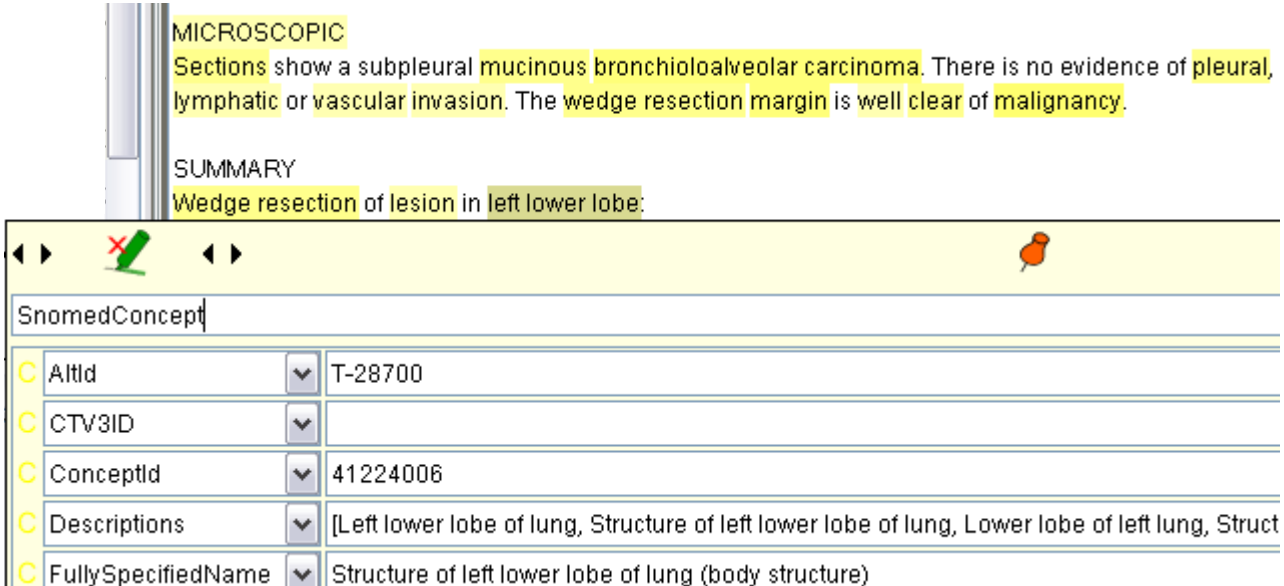
- *Sentence Splitter:*
  - Segment text into sentences using regular expressions
- *UMLS (MMTx) Annotator*:
  - Map strings in free text to closest concepts in UMLS
  - UMLS concepts were restricted to the SNOMED CT source



Pipeline Application

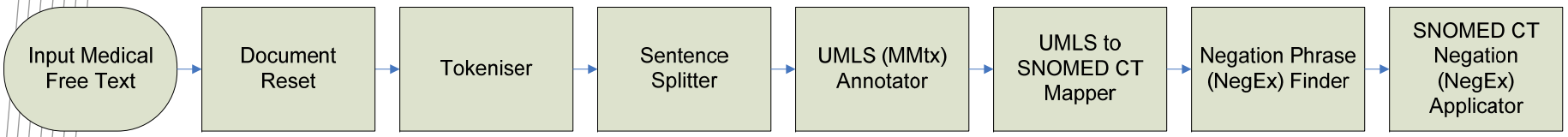


- *UMLS to SNOMED CT Mapper:*
  - UMLS concepts mapped to SNOMED CT using UMLS Metathesaurus database file
  - Only active concepts were retained
  - SNOMED ID tokens mapped to SNOMED CT concept



CSIRO

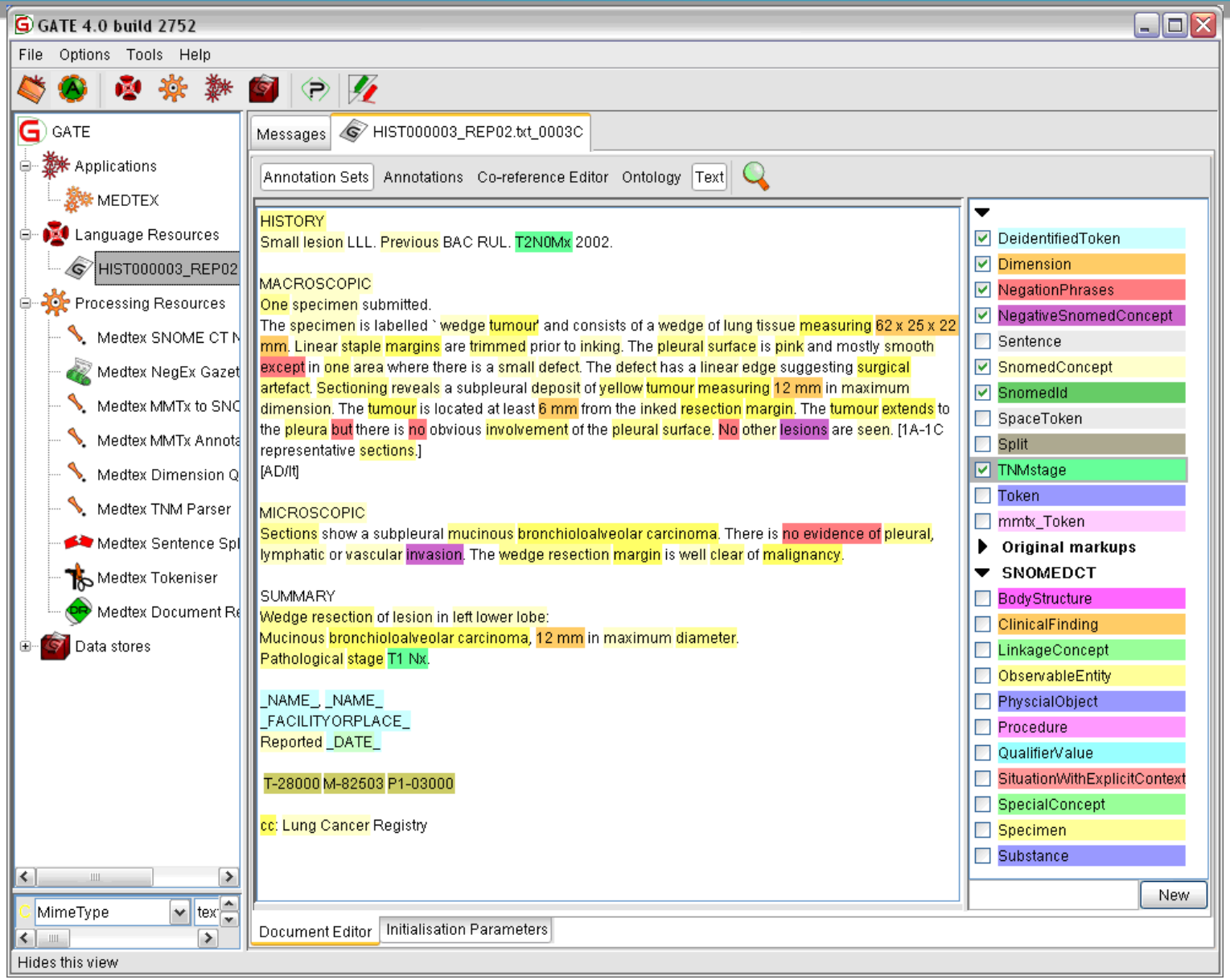
Pipeline Application



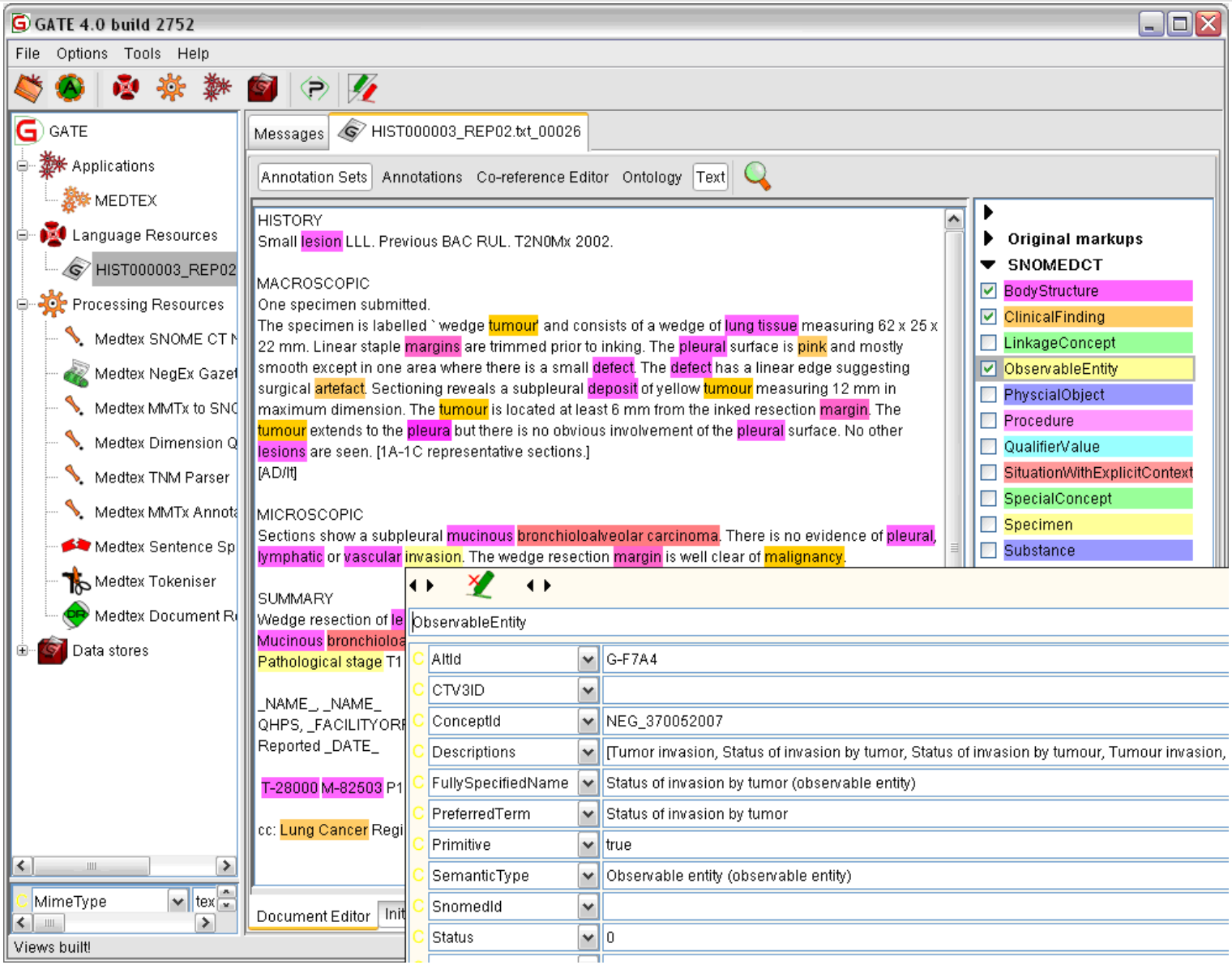
- *Negation Phrase (NegEx) Finder*
  - Finds common medical negation phrases
- *SNOMED CT Negation (NegEx) Applicator*
  - Tags “negated” concepts by associating negation phrases with neighbouring SNOMED CT “findings” or “disease” concepts
  - SNOMED CT sub-hierarchies considered for negation:
    - 404684003|clinical finding| (e.g. “decreased capillary fragility,” “diabetes mellitus,” “dyspnea.”)
    - 49755003|morphologically abnormal structure| (e.g. “wallerian degeneration,” “carcinoma”)
    - 363787002|observable entity| (e.g. “status of invasion by tumour”)
    - 272379006|event| (e.g. “exposure to toxin,” “accident caused by bench saw”)

CSIRO

Results


GATE 4.0 build 2752
File Options Tools Help
GATE
Applications
MEDTEX
Language Resources
HIST000003_REP02
Processing Resources
Medtex SNOME CT N
Medtex NegEx Gazet
Medtex MMTx to SNC
Medtex MMTx Annota
Medtex Dimension Q
Medtex TNM Parser
Medtex Sentence Spl
Medtex Tokeniser
Medtex Document Re
Data stores
MimeType
Messages
HIST000003_REP02.txt_0003C
Annotation Sets Annotations Co-reference Editor Ontology Text
HISTORY
Small lesion LLL. Previous BAC RUL. T2N0Mx 2002.
MACROSCOPIC
One specimen submitted.
The specimen is labelled `wedge tumour' and consists of a wedge of lung tissue measuring 62 x 25 x 22 mm. Linear staple margins are trimmed prior to inking. The pleural surface is pink and mostly smooth except in one area where there is a small defect. The defect has a linear edge suggesting surgical artefact. Sectioning reveals a subpleural deposit of yellow tumour measuring 12 mm in maximum dimension. The tumour is located at least 6 mm from the inked resection margin. The tumour extends to the pleura but there is no obvious involvement of the pleural surface. No other lesions are seen. [1A-1C representative sections.]
[AD/lt]
MICROSCOPIC
Sections show a subpleural mucinous bronchioloalveolar carcinoma. There is no evidence of pleural, lymphatic or vascular invasion. The wedge resection margin is well clear of malignancy.
SUMMARY
Wedge resection of lesion in left lower lobe:
Mucinous bronchioloalveolar carcinoma, 12 mm in maximum diameter.
Pathological stage T1 Nx.
NAME, _NAME_
FACILITYORPLACE
Reported _DATE_
T-28000 M-82503 P1-03000
cc: Lung Cancer Registry
DeidentifiedToken
Dimension
NegationPhrases
NegativeSnomedConcept
Sentence
SnomedConcept
SnomedId
SpaceToken
Split
TNMstage
Token
mmtx_Token
Original markups
SNOMEDCT
BodyStructure
ClinicalFinding
LinkageConcept
ObservableEntity
PhyscialObject
Procedure
QualifierValue
SituationWithExplicitContext
SpecialConcept
Specimen
Substance
New
Document Editor
Initialisation Parameters
Hides this view

Results


GATE 4.0 build 2752
File Options Tools Help
GATE
Applications
MEDTEX
Language Resources
HIST000003_REP02
Processing Resources
Medtex SNOME CT
Medtex NegEx Gazet
Medtex MMTx to SNC
Medtex Dimension Q
Medtex TNM Parser
Medtex MMTx Annota
Medtex Sentence Sp
Medtex Tokeniser
Medtex Document R
Data stores
MimeType
Views built!
Messages
HIST000003_REP02.txt_00026
Annotation Sets
Annotations
Co-reference Editor
Ontology
Text
HISTORY
Small lesion LLL. Previous BAC RUL. T2N0Mx 2002.
MACROSCOPIC
One specimen submitted.
The specimen is labelled ` wedge tumour' and consists of a wedge of lung tissue measuring 62 x 25 x 22 mm. Linear staple margins are trimmed prior to inking. The pleural surface is pink and mostly smooth except in one area where there is a small defect. The defect has a linear edge suggesting surgical artefact. Sectioning reveals a subpleural deposit of yellow tumour measuring 12 mm in maximum dimension. The tumour is located at least 6 mm from the inked resection margin. The tumour extends to the pleura but there is no obvious involvement of the pleural surface. No other lesions are seen. [1A-1C representative sections.]
[AD/lt]
MICROSCOPIC
Sections show a subpleural mucinous bronchioloalveolar carcinoma. There is no evidence of pleural, lymphatic or vascular invasion. The wedge resection margin is well clear of malignancy.
SUMMARY
Wedge resection of le
Mucinous bronchioloa
Pathological stage T1
NAME, _NAME_
QHPS, _FACILITYORI
Reported _DATE_
T-28000 M-82503 P1
cc: Lung Cancer Regi
Document Editor
Original markups
SNOMEDCT
BodyStructure
ClinicalFinding
LinkageConcept
ObservableEntity
PhyscialObject
Procedure
QualifierValue
SituationWithExplicitContext
SpecialConcept
Specimen
Substance
ObservableEntity
AltId G-F7A4
CTV3ID
ConceptId NEG_370052007
Descriptions [Tumor invasion, Status of invasion by tumor, Status of invasion by tumour, Tumour invasion,
FullySpecifiedName Status of invasion by tumor (observable entity)
PreferredTerm Status of invasion by tumor
Primitive true
SemanticType Observable entity (observable entity)
SnomedId
Status 0

Symbolic Rule-Based Post-Coordination

Subsumption relationships can be taken advantage of to retrieve *concepts which relate to the same disease, anatomy or finding*, or infer if *two descriptions relate to the same concept*.

Example "Lung Resection" Subsumption Test

Test if following post-coordinated expression template

<procedure> :
{260686004|method| = *<procedure.method>*
,405813007|procedure site - Direct| = *<topology>* }

is subsumed by:

119746007|lung excision|

Symbolic Rule-Based Post-Coordination

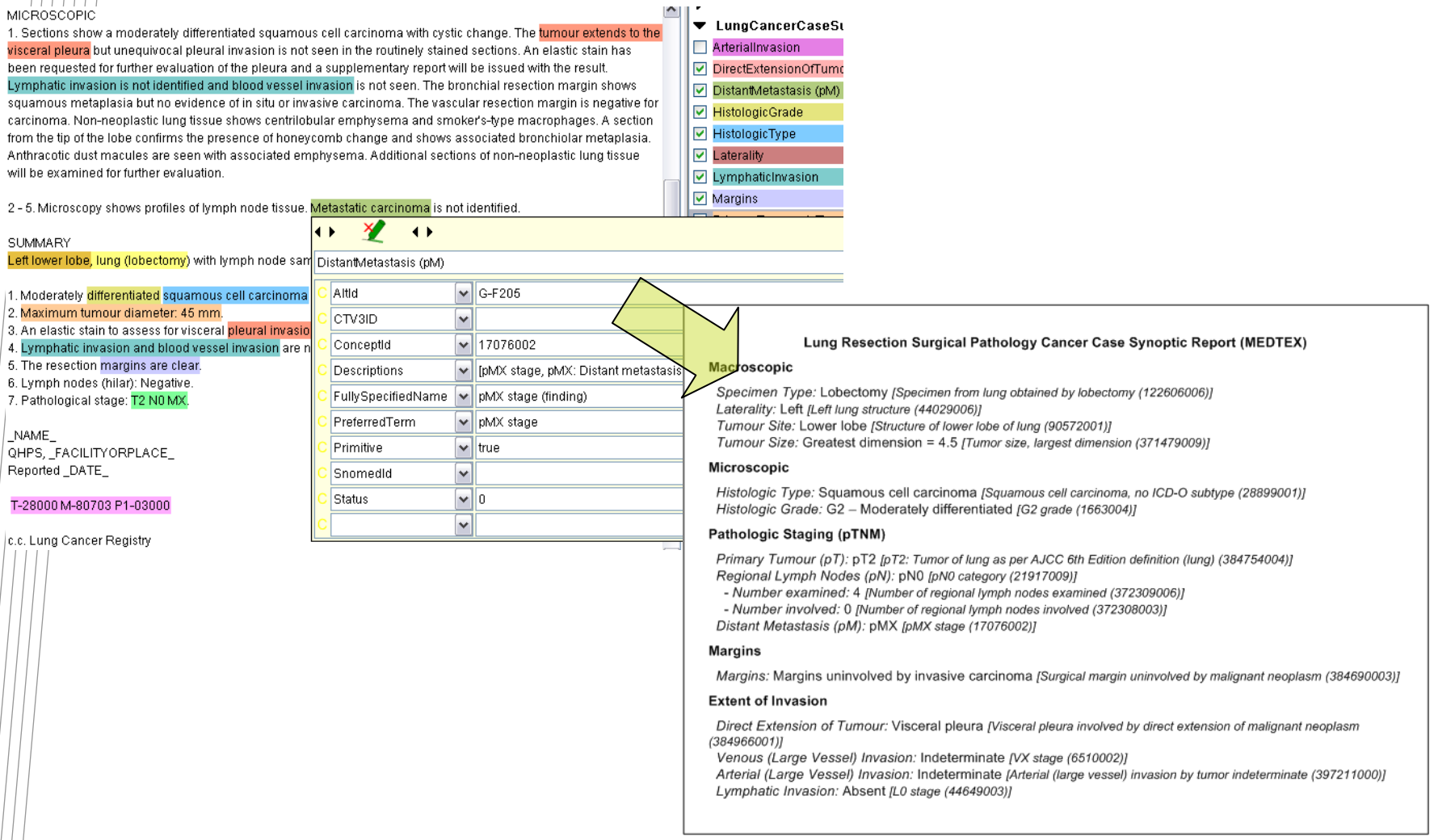


Figure 4. Synoptic factor extraction using SNOMED CT semantics

The Australian e-Health Research Centre
Michael Lawley
Principal Research Scientist

Phone: +61 7 3253 3609
Email: michael.lawley@csiro.au
Web: http://aehrc.com

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

Thank you

Contact Us
Phone: 1300 363 400 **or** +61 3 9545 2176
Email: Enquiries@csiro.au **Web:** www.csiro.au

CSIRO