

The SPECIALIST NLP Tools

Enhancing Synonym Features in the Lexical Tools

By: Dr. Chris J. Lu

[The Lexical Systems Group](#)

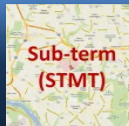
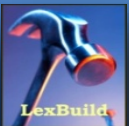
[NLM](#) – [LHNCBC](#) - [CGSB](#)

Oct., 2016

- Lexical Systems Group: <http://umlslex.nlm.nih.gov>
- The SPECIALIST NLP Tools: <http://specialist.nlm.nih.gov>

Table of Contents

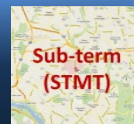
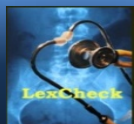
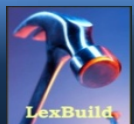
- Introduction
 - Natural Language Processing (NLP)
 - The SPECIALIST NLP Tools
- LexSynonyms
 - Introduction
 - Approach
 - Summary and Future Work
- Questions



Natural Language Processing (NLP)

- Map terms to concepts (meaning)
- Challenges: many to many mapping

Terms	Concepts	NLP
<ul style="list-style-type: none"> • cold • Cold Temperature • Cold Temperatures • Cold (Temperature) • Temperatures, Cold • Low temperature • low temperatures • ... 	<ul style="list-style-type: none"> • Cold Temperature C0009264 	<ul style="list-style-type: none"> • Concept mapping
<ul style="list-style-type: none"> • cold 	<ul style="list-style-type: none"> • Cold Temperature C0009264 • Common Cold C0009443 • Cold Therapy C0010412 • Cold Sensation C0234192 • ... 	<ul style="list-style-type: none"> • WSD (Word Sense Disambiguation)



NLP – Concept Mapping

➤ Normalization:

- A term might have a great deal of lexical variations, such as inflectional variants, spelling variants, derivations, abbreviations (expansions), cases, ASCII conversion, etc.
- Normalize different forms of a concept to a same form

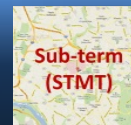
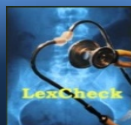
➤ Query Expansion:

- Expand a term to its equal terms, such as subterm substitution of synonyms, derivational variants, spelling variants, abbreviations, etc.
- To increase recall

➤ POS tagger:

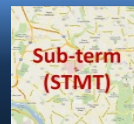
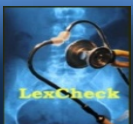
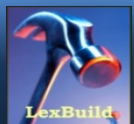
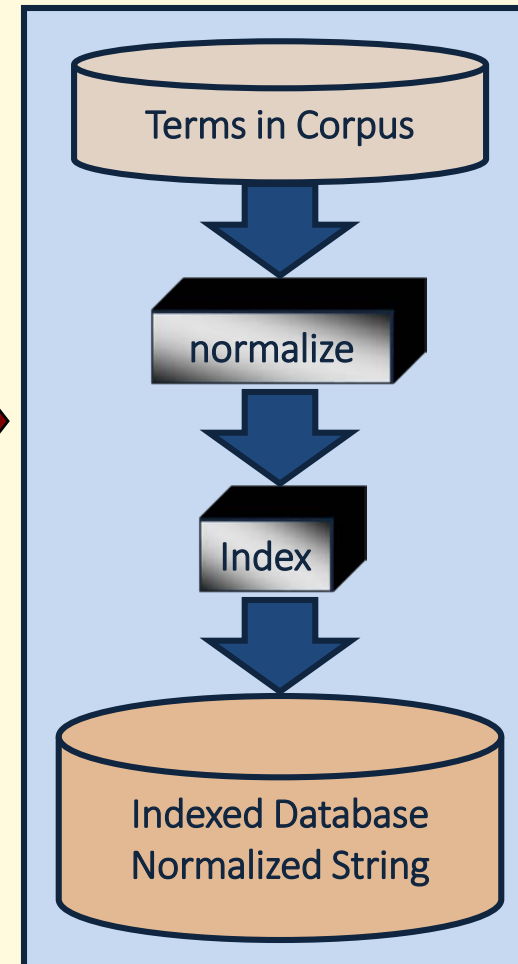
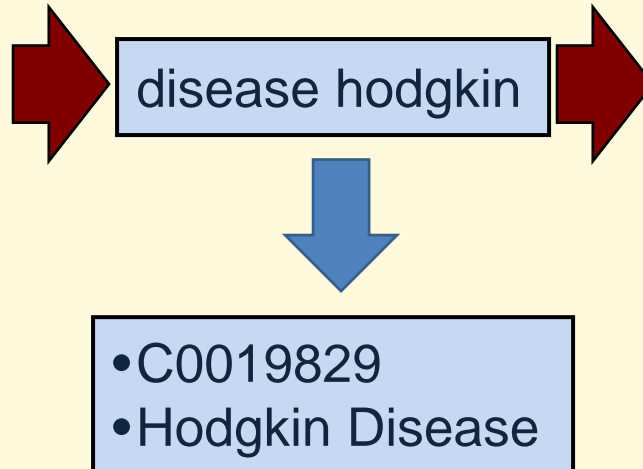
- Assign part of speech to a single word or multiword in a text
- To increase precision

➤ Others...

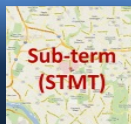
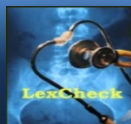
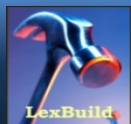
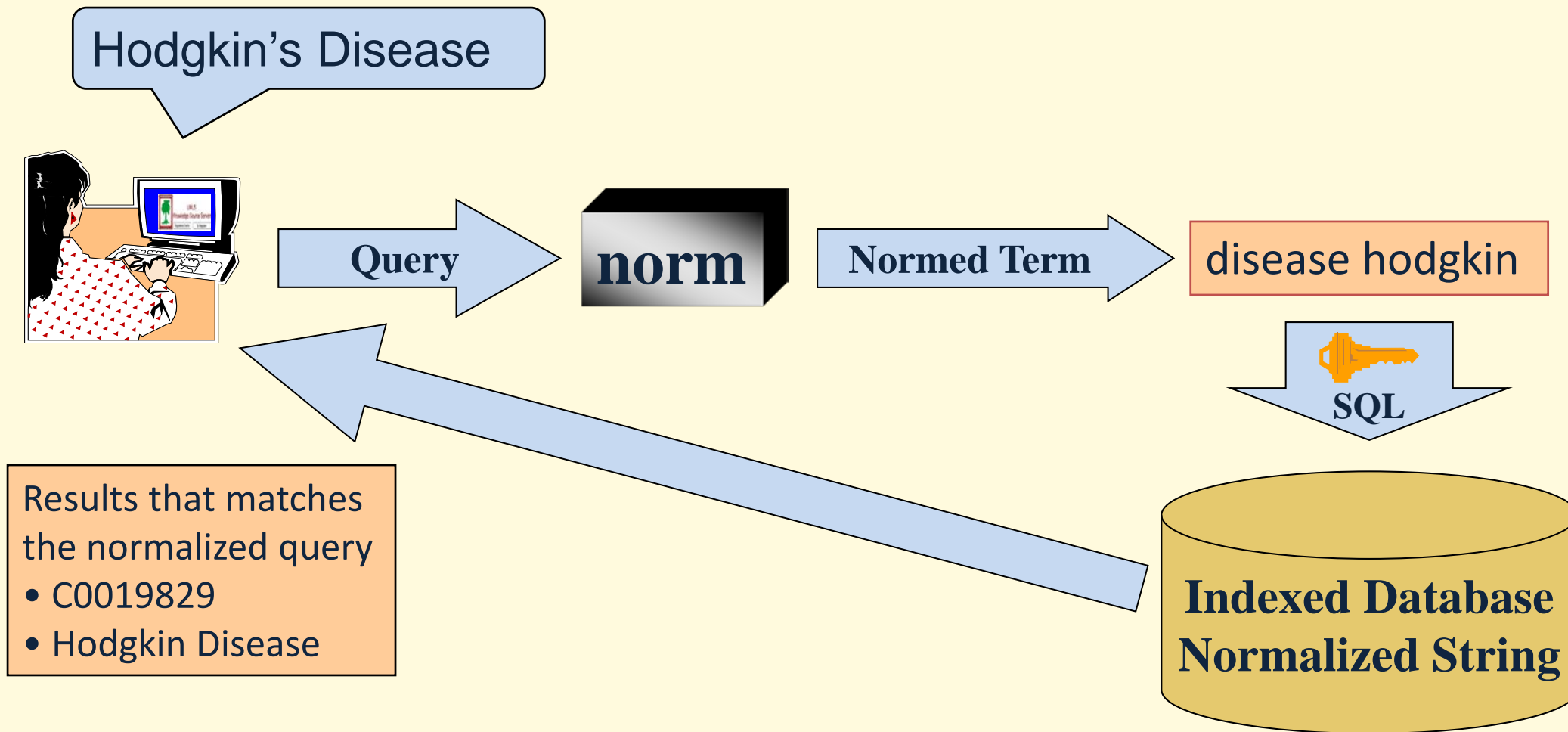


NLP – Norm (Lexical Variations)

- Hodgkin Disease
- HODGKINS DISEASE
- Hodgkin's Disease
- Disease, Hodgkin's
- HODGKIN'S DISEASE
- Hodgkin's disease
- Hodgkins Disease
- Hodgkin's disease NOS
- Hodgkin's disease, NOS
- Disease, Hodgkins
- Diseases, Hodgkins
- Hodgkins Diseases
- Hodgkins disease
- hodgekin's disease
- Disease;Hodgkins
- Disease, Hodgkin
- ...



NLP – Norm (Cont.)

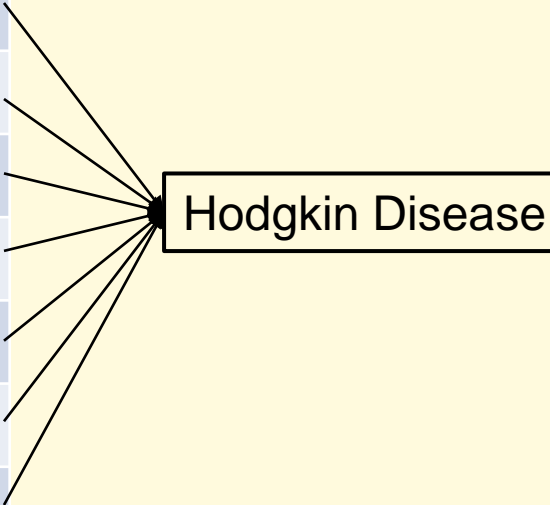


Normalized Patterns (in UMLS)

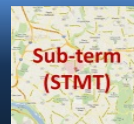
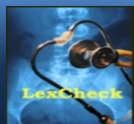
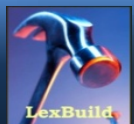
Example: Hodgkin Disease (C0019829)

- There are 139 terms associated with this concept
- Similar normalized patterns

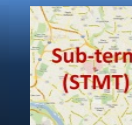
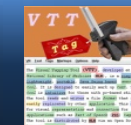
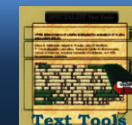
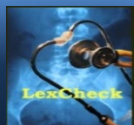
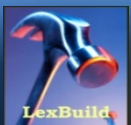
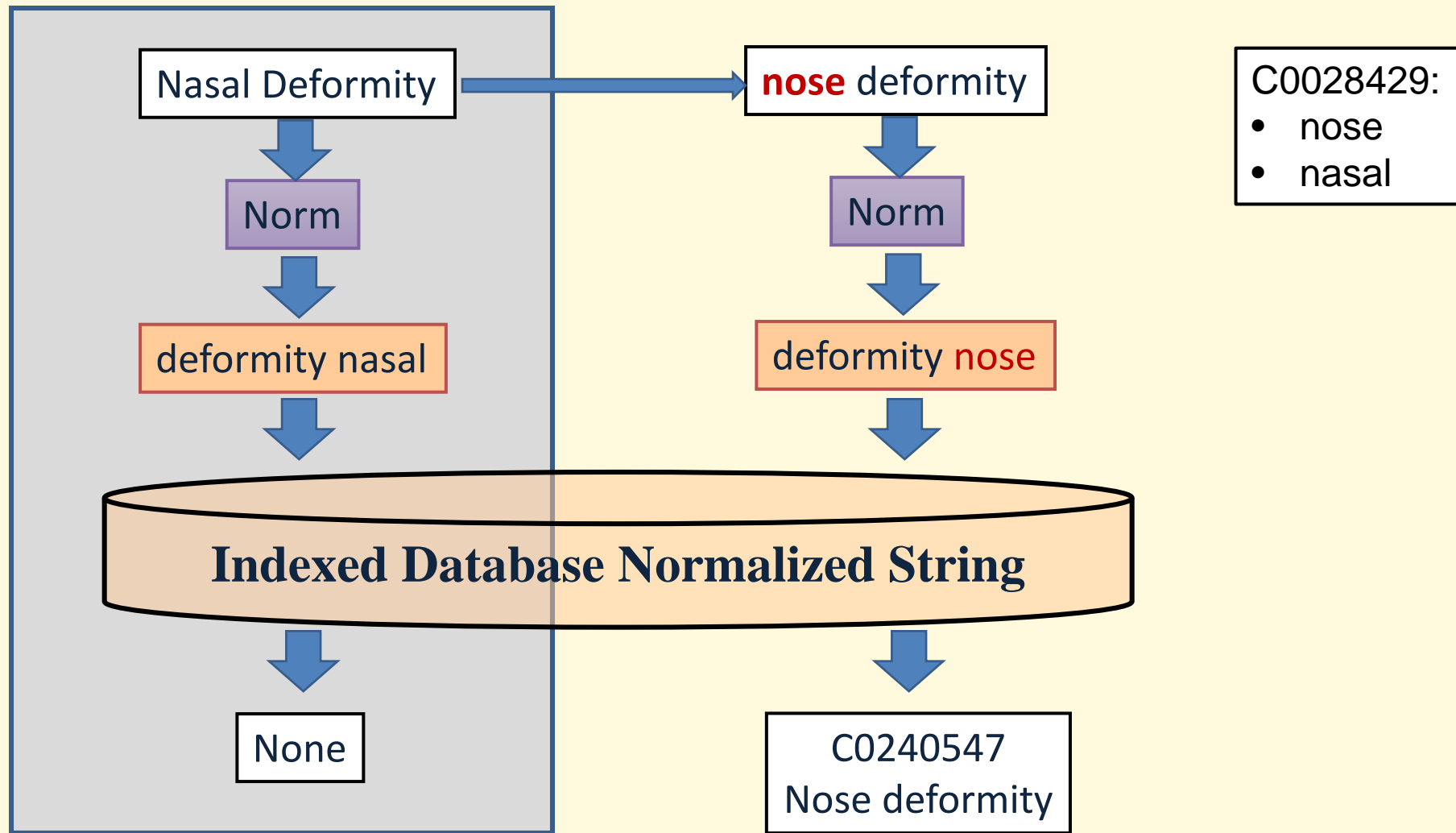
Pattern String	No. of Terms
hodgkin disease	32
hodgkin granuloma	22
hodgkin lymphoma	23
hodgkin paragranuloma	14
hodgkin sarcoma	15
...	...
Sub-total	139



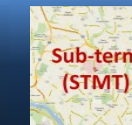
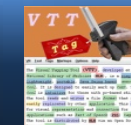
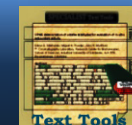
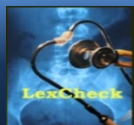
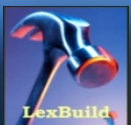
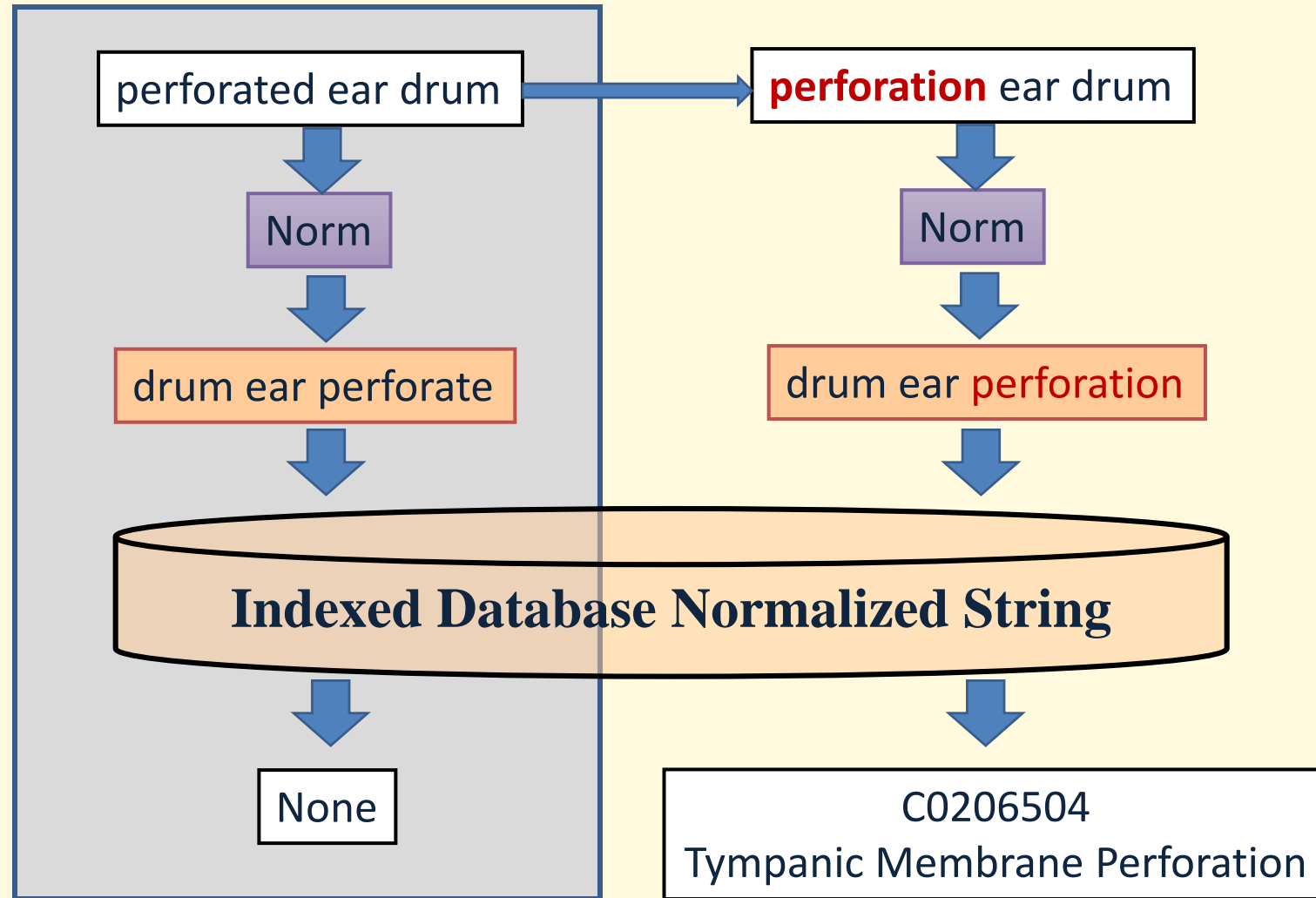
A diagram consisting of a rectangular box labeled "Hodgkin Disease" with a black border. Seven lines radiate from the left side of this box, pointing to the right side of the table above. The lines originate from the right edge of the table rows corresponding to the pattern strings: "hodgkin disease", "hodgkin granuloma", "hodgkin lymphoma", "hodgkin paragranuloma", "hodgkin sarcoma", "...", and "Sub-total".



NLP – Query Expansion (Synonym)

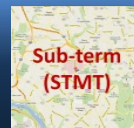
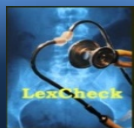
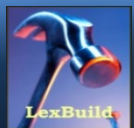


NLP – Query Expansion (derivation)



NLP – Query Expansion

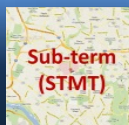
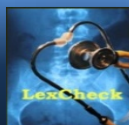
- Used to increase recall (find more relevant doc)
- Sub-term substitution thesauri:
 - Lexical variants (morphology)
 - Semantic related (synonym)
 - Statistical co-occurrence: heart rate, heart disease, heart failure, heart transplantation, heart surgery, heart transplant, heart valve, etc.
- Integrated with term frequency-inverse document frequency (Tf-idf) to assist in ranking (precision dropped)



Lexical Variants from the Lexicon and Lexical Tools



```
{base=generalise  
spelling_variant=generalize - - - - -> spelling variant  
entry=E0029526  
  cat=verb - - - - -> part of speech  
  variants=reg - - - - -> inflectional variant  
  intran  
  tran=np  
  tran=pphr(from,np) - - - - -> chunker  
  tran=pphr(to,np)  
  nominalization=generalisation|noun|E0029525 - - - - -> derivational variant, synonym  
}
```



Generated Lexical Variants

LexRecord: E0029526|generalise|verb

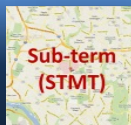
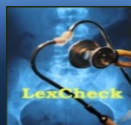
- POS: verb
- citation: generalise
- spVar: generalize
- inflVars: generalises, generalised, generalising
- nominalization: generalisation, generalization
- Abbreviation/acronym: n/a

Derivational variants:

- suffixD: generalis**ation**, generaliz**ation**, generalis**able**
- prefixD: **over**generalise, **over**-generalise

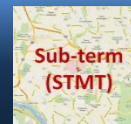
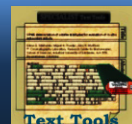
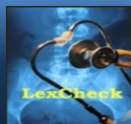
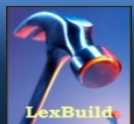
Synonyms: generalize

Fruitful Variants: generalisability, generalisable, generalisation, generalisations, generalised, generalises, generalising, generalizability, generalizable, generalization, generalizations, generalize, generalized, generalizer, generalizers, generalizes, generalizing, overgeneralize, etc.



Lexical Tools – Flow Components (62)

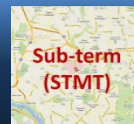
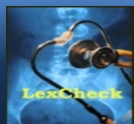
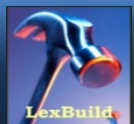
Lexicon Related – Data (32)	Non-Lexicon related – Algorithm (30)
Inflection (10): b, B, Bn, l, ici, is, L, Ln, Lp, si,	Unicode operation (10): q, q0, q1, q2, q3, q4, q5, q6, q7, q8
Derivation (3): d, dc, R	Tokenizer (3): c, ca, ch
Acronym or abbreviation (3): a, A, fa	Punctuation operation (3): o, p, P
Spelling variant (2): e, s	Lowercase (1): l
Lexicon mapping (3): An, E, f, fp	Metaphone (1): m
Synonym (2): y, r	Remove parenthetic plural forms (1): rs
Nominalization (1): nom	Strip stop word (1): t
Citation (1): Ct	Remove genitive (1): g
Fruitful variant (4): G, Ge, Gn, V	No operation (1): n
Normalization (2): N, N3,	...



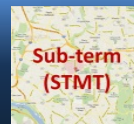
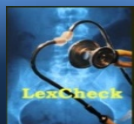
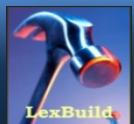
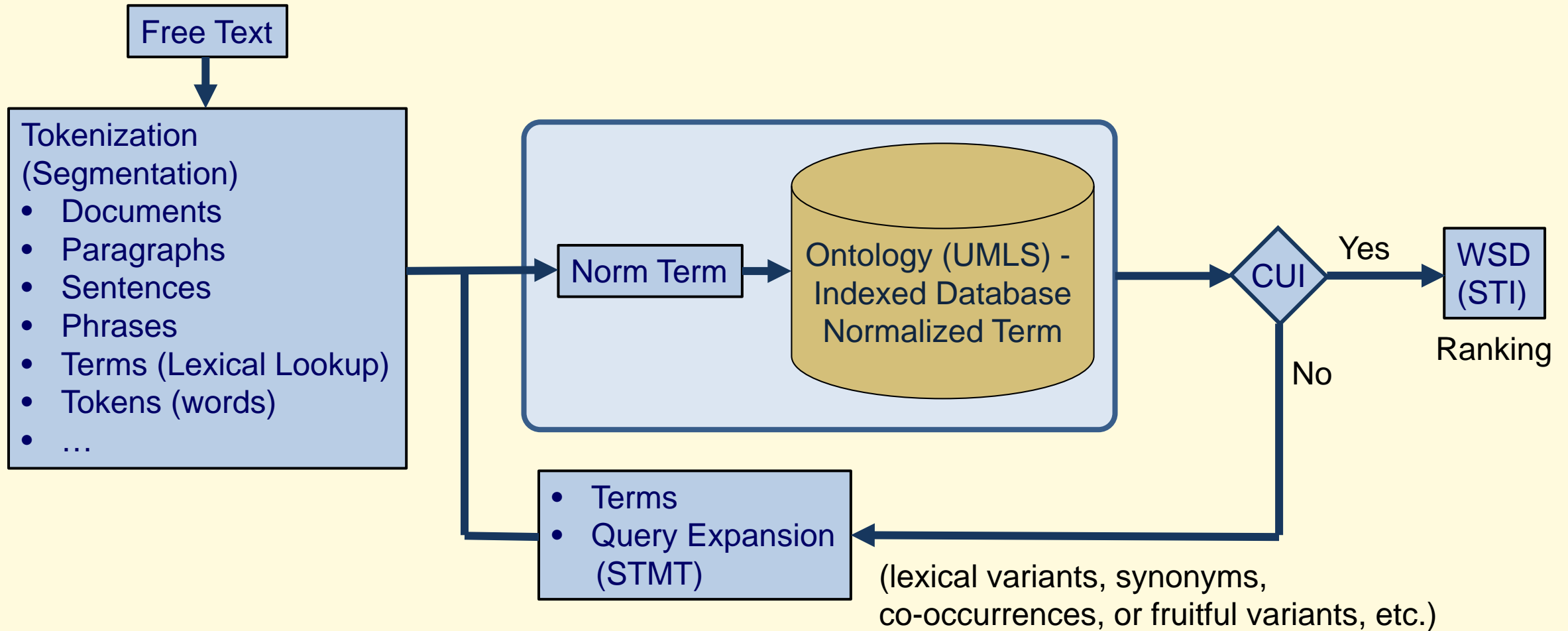
LVG - Lexical Variants Generation

➤ To increase recall & precision

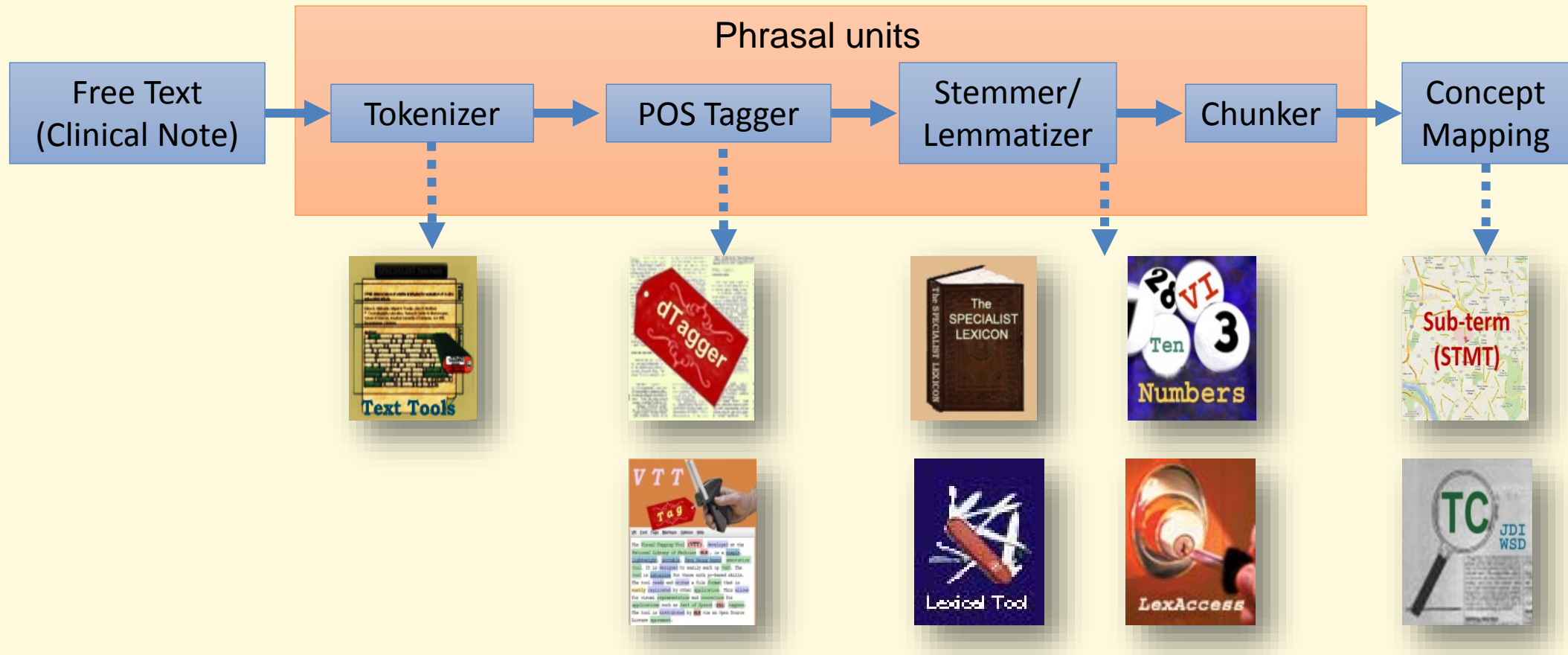
	Recall - Query Expansion	Precision - POS Tagging
Inputs	perforated ear drum	saw
UMLS-CUI	None	<ul style="list-style-type: none"> • C1947903 verb see • C0183089 noun saw (device)
Process	perforation ear drum	noun
UMLS-CUI	C0206504	<ul style="list-style-type: none"> • C0183089
Preferred Term	Tympanic Membrane Perforation	saw (device)



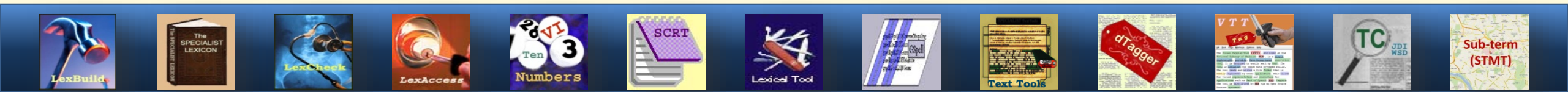
NLP – Concept Mapping Model



The SPECIALIST NLP Tools



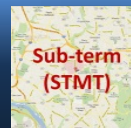
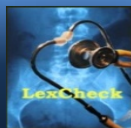
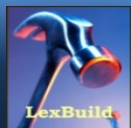
- Lexical Systems Group: <http://umlslex.nlm.nih.gov>
- The SPECIALIST NLP Tools: <http://specialist.nlm.nih.gov>



The SPECIALIST NLP Tools by LSG



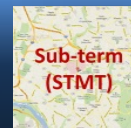
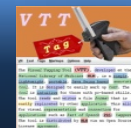
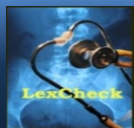
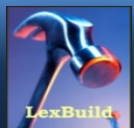
- Lexical Systems Group: <http://umlslex.nlm.nih.gov>
- The SPECIALIST NLP Tools: <http://specialist.nlm.nih.gov>



Synonyms

Synonyms are words (terms) that have the same or very similar *meaning (concept)*

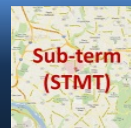
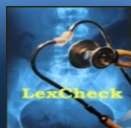
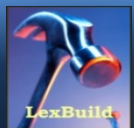
- *Strings with same CUI in the UMLS Metathesaurus (UMLS Synonyms)*
- The 2016AA Metathesaurus contains more than 3.25 million concepts (CUIs) and nearly 13 million unique concept names (AUIs) from over 190 source vocabularies.
- UMLS synonyms: terms have same Concept Unique Identifier (CUI)
 - C0018592 | Happiness (22):
joy, enjoy, happy, joyful, enjoyment, happiness, happy mood, mood happy, high spirits, good mood, in good spirits, bright in mood, cheerful mood, affect happy, enjoyed, feeling of joy, etc..
 - C0018681 | Headache (43):
headache, head pain, pain in head, cephalalgia, cephalodynia, cranial pain, head pain cephalgia, ...



UMLS Synonyms

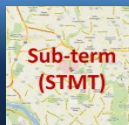
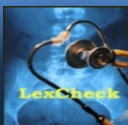
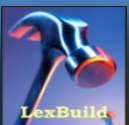
Synonyms are words (terms) that have the same *concept (CUI)*

- *Strings with same CUI in the UMLS Metathesaurus (UMLS Synonyms)*
- Similarity & relatedness
- Meaning changes a lot based on domain (one term have multiple concepts)
- Over-generated - word or entire phrase
 - Synonym strings (UMLS Strings) – not necessary real (multi)words
- No POS



Synonym Types

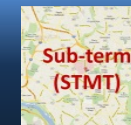
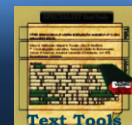
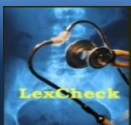
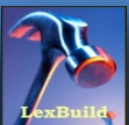
- Cognitive synonym:
 - less difference
 - greater interchangeability (not context-sensitive)
 - can be represented as a synonym pair (sPair)
- Near-synonym:
 - greater difference
 - less interchangeability
 - specific use, can't used in generic case



Properties of sPairs

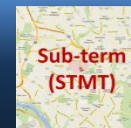
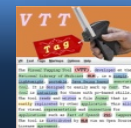
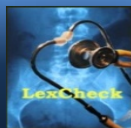
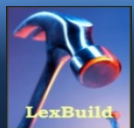
- Commutativity: $(x = y) \rightarrow (y = x)$
 - bi-directional
 - joy | noun | enjoy | verb \rightarrow enjoy | verb | joy | noun

- Transitivity: $((x = y) \text{ and } (y = z)) \rightarrow (x = z)$
 - enjoy | verb | joy | noun
 - joy | noun | happy | adj
 - >
 - recursive
 - enjoy | verb \rightarrow joy | noun \rightarrow happy | adj



The Synonyms in Lexical Tools (2016-)

- Developed in early 90's
- The original idea is to provide synonyms that are not in the UMLS Metathesaurus
 - not a complete data set
- Manually updated by user's requests (static):
 - 2004 (5,056) -> 2016 (5,198)
 - Only 142 sPairs were added since 2004
 - Need an automatic/systematic way to generate synonyms
- Not necessary good sPairs
- 6 associated flow components (10%): G, Ge, Gn, r, v, y



Recursive Issue in Lexical Synonyms.2016

➤ Transitivity: $((x \sim y) \text{ and } (y \sim z)) \rightarrow (x \sim z)$

- Lexical Tools (2006-):

- C0036658 (esthesia)

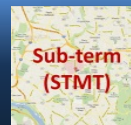
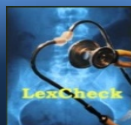
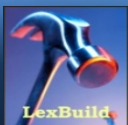
sensation|noun ↔ esthetic|adj

- C0014901 (esthetics)

esthetic|adj ↔ beauty|noun

- How about recursive synonym?

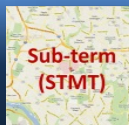
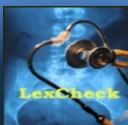
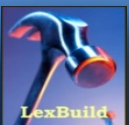
sensation|noun ↔ esthetic|adj ↔ beauty|noun



Objectives

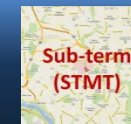
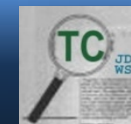
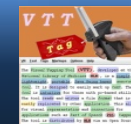
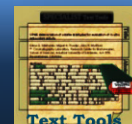
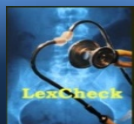
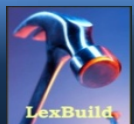
- To establish a system to:
 - generate a standalone synonym thesaurus
 - include all synonymous terms in Lexicon (LexSynonyms)
 - grow with the SPECIALIST Lexicon

- improve NLP performance
 - by resolving issues of using UMLS synonyms



Broader Issues – Near Synonyms

CUI	Preferred Term	synonym	Explanation
C0001613	Adrenal Cortex	cortical	The adjective cortical can refer to any of several types of cortex & so does not have synonymy with “adrenal cortex”
C0032639	Pontine structure	metencephalon	The metencephalon, per m-w.com includes the cerebellum and pons, and is different from the pons
C0001575	Uterine adnexae structure	adnexa	There are several types of adnexa, such as eye adnexa, adnexa of skin, etc.
C0000936	Visual Accommodation	accommodation	There are other accommodations.



Metencephalon & Pontine Structure (Pons)

Hinbrain: *Metencephalon*

20

b) metencephalon

▣ pons

- Contains pneumotaxic centre which fine tunes breathing rate
- Relays information between cerebellum and cerebrum

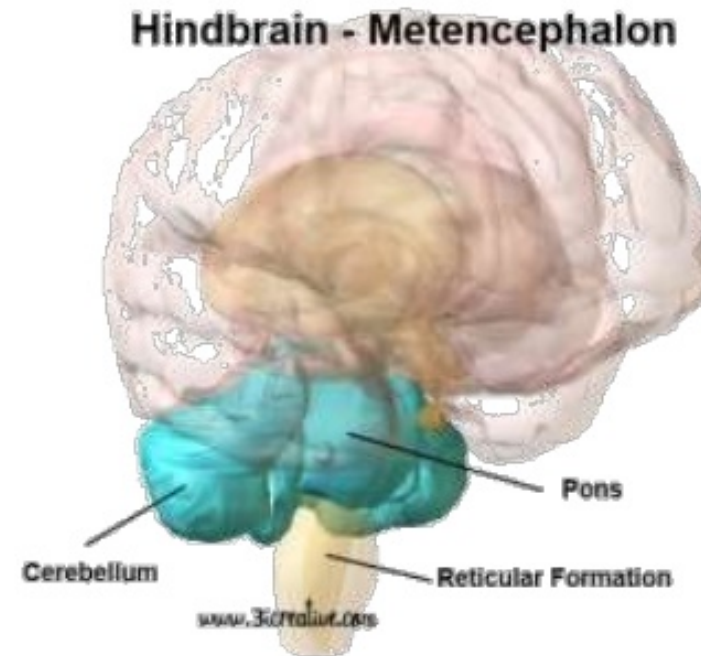
▣ cerebellum

- Feedback center for execution of motor movements
- Controls posture and balance

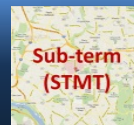
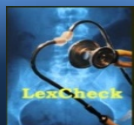
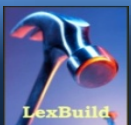
▣ reticular formation

- Nuclei diffusely located through the brainstem*
- Regulates wakefulness and muscle tone

Hindbrain - Metencephalon

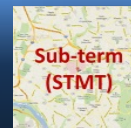
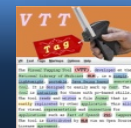
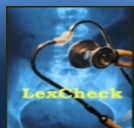
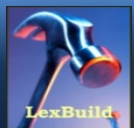


*the term "brainstem" refers to the medulla oblongata, pons, and the midbrain



Distinct Issues – Similar but Different

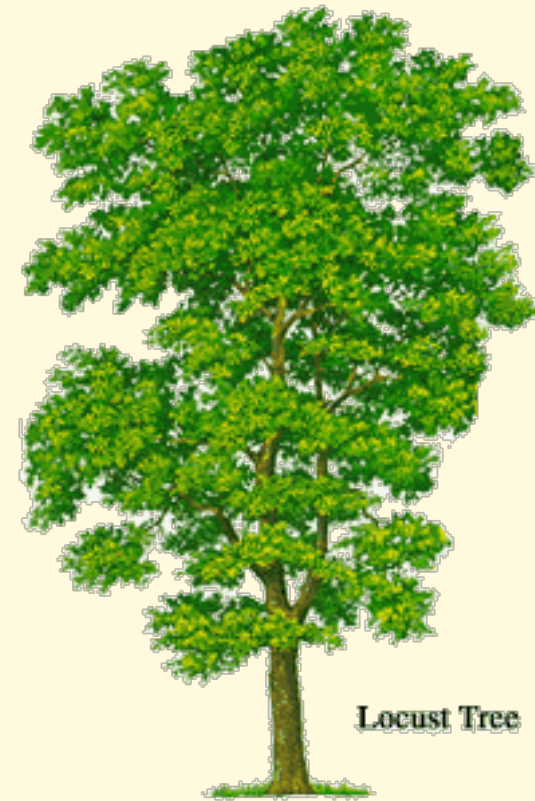
CUI	Preferred Term	synonym	Explanation
C0000741	Abducens nerve structure	abductor digiti minimi	The abductor digiti minimi is a muscle, not a nerve.
C0003864	Arthritis	arthritide	Per Dorland's an arthride is "any skin eruption of arthritic or gouty origin.
C0005400	Bile duct structure	choledochitis	Choledochitis is a condition of the common bile duct, not structure.
C0000869	Acacia	locust tree	Though both the acacia & locust tree are members of Leguminosae (pea, bean), they do seem to refer to different trees.
C0003353	Antigua	Anguilla	The islands of Antigua & Anguilla are both in the West Indies, but are not the same place.



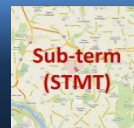
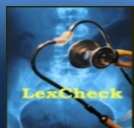
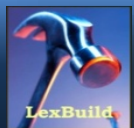
Acacia & Locust tree



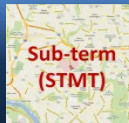
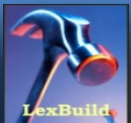
Acacia



Locust Tree

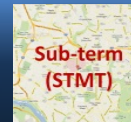
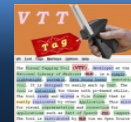
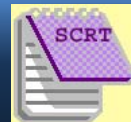
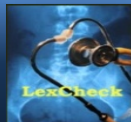


Anguilla & Antigua



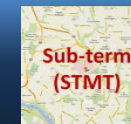
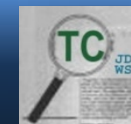
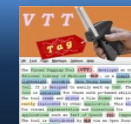
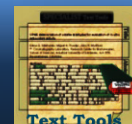
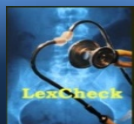
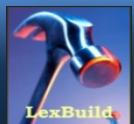
Acronym/Abbreviation Issues – Precision

CUI	Preferred Term	synonym
C0001175	Acquired Immunodeficiency Syndrome	sida
C0001857	AIDS related complex	arc
C0003023	Angola	ago
C3714936	Non-Compliant ADaM Datasets Domain	ax



POS Issues – Meaning Shift

CUI	Preferred Term	synonym	Explanation
C0001774	Agaricales	Mushroom	The verb (to) mushroom does not refer to Agaricales while the noun is a synonym
C0003459	Anura	frog	The verb (to) frog does not refer to Anura, while the noun is a synonym
C0003842	Arteries	arterial	The noun arterial refers to roads, not circulatory anatomy, unlike the adjective arterial.
C0004063	Assault	mug	The verb mug does refer to assault, but the noun mug does not.



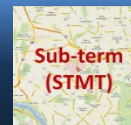
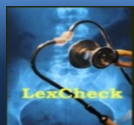
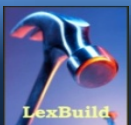
Recursive Issues – Multiple Concepts

- Multiple CUIs (transitivity?)
- Example (cold):

CUIs	Synonym
C0009443 common cold	<ul style="list-style-type: none"> • cold • coryza • acute coryza • common cold
C0009264 cold temperature	<ul style="list-style-type: none"> • cold • low temperature • low-temperature • lowtemperature
C0234192 cold sensation	<ul style="list-style-type: none"> • cold • psychroesthesia
...	• ...

- common cold|cold
- cold|cold temperature
- cold|cold sensation

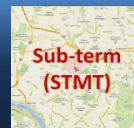
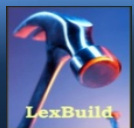
=> common cold|cold|cold temperature ?
=> common cold|cold|cold sensation?



Recursive Issue 2 – Endless loop

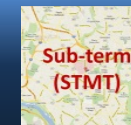
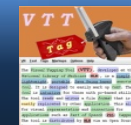
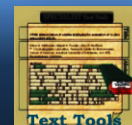
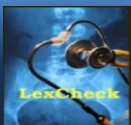
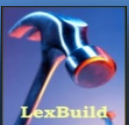
➤ Example – cold blooded animal

Synonyms	cold temperature	cold therapy	common cold	cold sensation
1-G substitution	cold temperature	cold therapy	common cold	cold sensation
2-G substitution	cold temperature temperature cold therapy temperature common cold temperature cold sensation temperature	cold temperature therapy cold therapy therapy common cold therapy cold sensation therapy	common cold temperature common cold therapy common common cold common cold sensation	...
...				



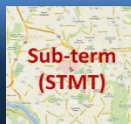
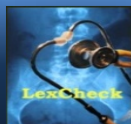
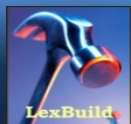
LexSynononyms – Synonyms in Lexicon

- Requirements (sClass):
 - All synonymous terms in the Lexicon
 - Bi-directional (commutativity) - interchangeable sPair in NLP
 - Recursive (transitivity) - use in NLP to increase Recall, without dropping precision too much
- Resolve all above issues
 - Broader issues
 - Distinct issues
 - Acronym/abbreviation issues
 - POS issues
 - Recursive issues



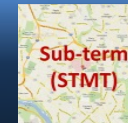
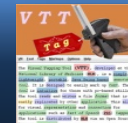
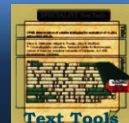
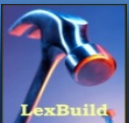
Synonym Class (sClass) on UMLS Synonyms

- English terms from MRCONSO.RRF with same CUI
- Exclude chemicals & drugs
 - use MRSTY.RRF to map CUI to STI
 - filter out disallowed STI in SemGroups.filter.txt
- In Lexicon with inflection is base and POS of adj, noun, or verb
- Remove acronyms/abbreviations => it drops precision
- Remove spVars => add them in post-process
- Remove nominalization => add them in post-process
- Remove singleton sClass (1 single candidates)
- Manually verify



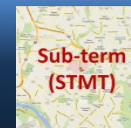
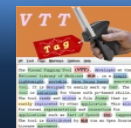
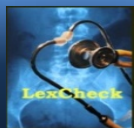
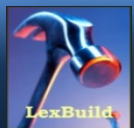
Synonym Class Example

```
#SYNONYM_CLASS|C0003842|Arteries  
128|E0010481|arteria|Y  
128|E0010531|artery|Y  
128|E0694191|arterial|N  
1|E0010482|arterial|Y  
#SYNONYM_CLASS|C0004063|Assault  
1024|E0041250|mug|Y  
128|E0010822|assault|Y  
128|E0041249|mug|N  
...
```



Approach – Synonym Sources

- NLP Projects, Lexical Tools 2016- (LVG), STMT, UMLS Core, ...
 - legacy data
 - can be automatically retrieved
 - Verify and add POS
- Lexicon (EUI)
 - nominalizations are synonyms
 - can be retrieved from Lexicon automatically
 - associated EUIs are preserved
- Metathesaurus (CUI)
 - retrieve synonym candidates (sClass)
 - tag synonym candidate list
 - associated CUIs are preserved

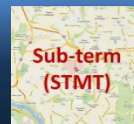
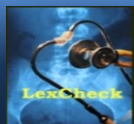
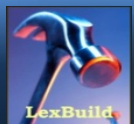


LexSynonym Data File – sClass

```
...  
#SYNONYM_CLASS|C0011065|Cessation of life  
128|E0020918|death|Y  
1|E0020877|dead|Y  
1|E0020990|deceased|Y  
1|E0022536|die|  
...
```

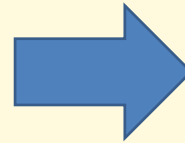
```
{base=dead  
entry=E0020877  
  cat=adj  
  variants=inv  
  variants=reg  
  position=attrib(1)  
  position=attrib(3)  
  position=pred  
  stative  
  nominalization=deadness|noun|E0020885  
}
```

```
{base=death  
entry=E0020918  
  cat=noun  
  variants=reg  
  variants=uncount  
  compl=pphr(of,np)  
  compl=pphr(from,np)  
  nominalization_of=die|verb|E0022536  
}
```

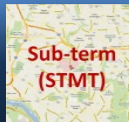
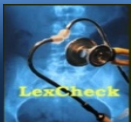


LexSynonym Data File

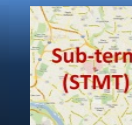
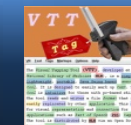
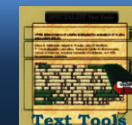
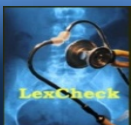
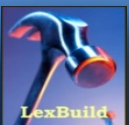
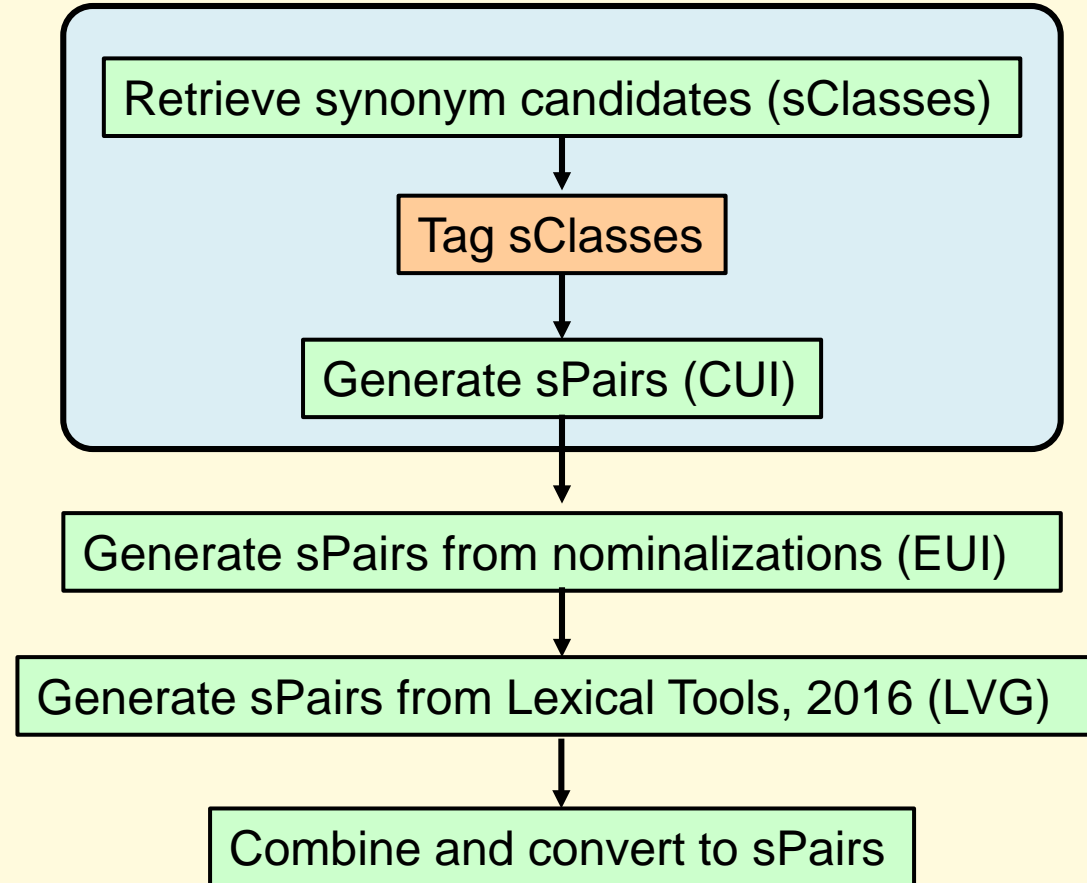
...
#SYNONYM_CLASS | C0011065 | Cessation of life
128 | E0020918 | death | Y
1 | E0020877 | dead | Y
1 | E0020990 | deceased | Y
1024 | E0022536 | die | nom
128 | E0020885 | deadnes | nom
...



...
deadness | 128 | dead | 1 | C0011065
deadness | 128 | death | 128 | C0011065
deadness | 128 | deceased | 1 | C0011065
deadness | 128 | die | 1024 | C0011065
dead | 1 | deadness | 128 | C0011065
dead | 1 | death | 128 | C0011065
dead | 1 | deceased | 1 | C0011065
dead | 1 | die | 1024 | C0011065
death | 128 | deadness | 128 | C0011065
death | 128 | dead | 1 | C0011065
death | 128 | deceased | 1 | C0011065
death | 128 | die | 1024 | C0011065
deceased | 1 | deadness | 128 | C0011065
deceased | 1 | dead | 1 | C0011065
deceased | 1 | death | 128 | C0011065
deceased | 1 | die | 1024 | C0011065
die | 1024 | deadness | 128 | C0011065
die | 1024 | dead | 1 | C0011065
die | 1024 | death | 128 | C0011065
die | 1024 | deceased | 1 | C0011065
...



LexSynonym Generation



Lexical Tools – Synonym Flow

- Software Changes:
 - Include POS and the source information in synonym database

- Example:

```
shell> lvg -f:y -m
```

```
die
```

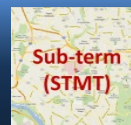
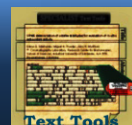
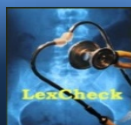
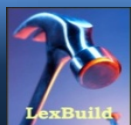
```
die|dead|1|1|y|1|FACT|die|die|verb|dead|adj|C0011065|
```

```
die|deadness|128|1|y|1|FACT|die|die|verb|deadness|noun|C0011065|
```

```
die|death|128|1|y|1|FACT|die|die|verb|death|noun|C0011065|
```

```
die|deceased|1|1|y|1|FACT|die|die|verb|deceased|adj|C0011065|
```

```
die|expire|1024|1|y|1|FACT|die|die|verb|expire|verb|NLP_LVG|
```



Lexical Tools – Synonyms Flow Options

- Synonym source restriction options (-ks):
 - C (CUI), E (EUI), N (NLP), CE, CN, EN, CEN.

- Example:

```
shell> lvg -f:y -m -ks:C
```

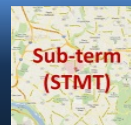
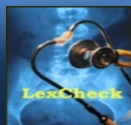
```
die
```

```
die|dead|1|1|y|1|FACT|die|die|verb|dead|adj|C0011065|
```

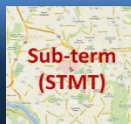
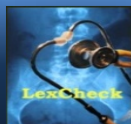
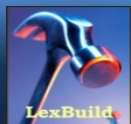
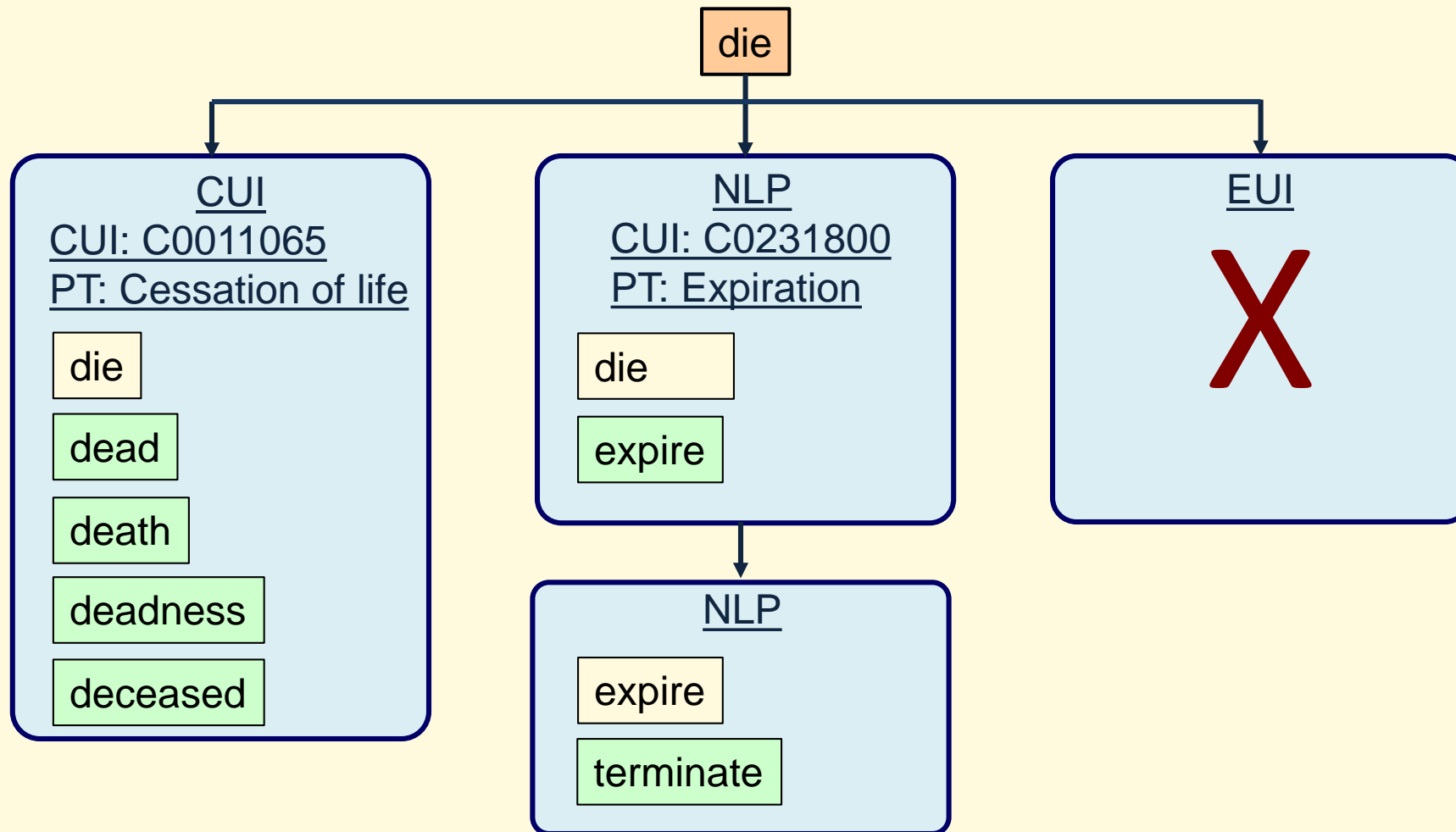
```
die|deadness|128|1|y|1|FACT|die|die|verb|deadness|noun|C0011065|
```

```
die|death|128|1|y|1|FACT|die|die|verb|death|noun|C0011065|
```

```
die|deceased|1|1|y|1|FACT|die|die|verb|deceased|adj|C0011065|
```



Lexical Tools – Recursive Synonyms



Lexical Tools – Recursive Synonym Flow

➤ Software Changes:

- must have the same type of source
- If the source is CUI: only synonyms from the same CUI are used
- If the source is EUI: all synonyms with EUI source are used
- If the source is NLP: synonyms from same NLP source are used

➤ Example:

```
shell> lvg -f:y -m
```

```
die
```

```
die|dead|1|1|r|2|FACT|die|verb|dead|adj|C0011065|y|
```

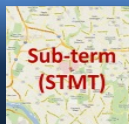
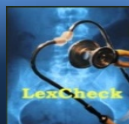
```
die|deadness|128|1|r|2|FACT|die|verb|deadness|noun|C0011065|y|
```

```
die|death|128|1|r|2|FACT|die|verb|death|noun|C0011065|y|
```

```
die|deceased|1|1|r|2|FACT|die|verb|deceased|adj|C0011065|y|
```

```
die|expire|1024|1|r|2|FACT|die|verb|expire|verb|NLP_LVG|y|
```

```
die|terminate|1024|1|r|2|FACT|expire|verb|terminate|verb|NLP_LVG|yy|
```



Results

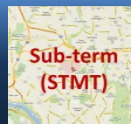
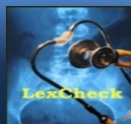
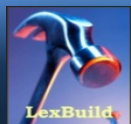
- 2017 release:
 - 2016AB Metathesaur, 2016 Lexicon

	Total	Tagged	Completion (%)
sClass	22,779	7,686	33.74%
Synonyms	80,913	29,990	37.06%

- Synonym stats:

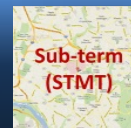
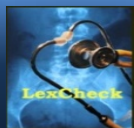
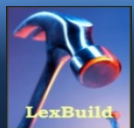
Year	CUI	EUI	NLP	Total
2016	0 (0%)	0 (0%)	5,198 (100%)	5,198
2017	118,468 (62%)	67,584 (35%)	4,792 (3%)	190,844

36.71 growth



Summary

Goals	Check	Notes
Standalone synonym database	Yes	
All synonymous terms in the Lexicon	Yes	~ 1/3 completed
Grows with the SPECIALIST Lexicon	Yes	
Over-generated issues	Yes	Must be terms in the Lexicon
Broader issues	Yes	Done in tagging
Distinct issues	Yes	Done in tagging
Acronym/abbreviation issues	Yes	Removed in sClass
POS issues	Yes	Provide POS in sClass
Recursive issues	Yes	Provide source in sClass
Improve NLP performance	Theoretically	To be tested

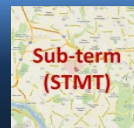
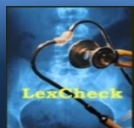
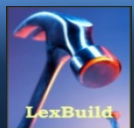


Future Work

- Complete tagging on all sClasses
 - 2017 release: 2016AB Metathesaurus, 2016 Lexicon

	Total	Tagged	Completion (%)
sClass	22,779	7,686	33.74%
Synonyms	80,913	29,990	37.06%

- Updated annually on Lexicon and Lexical Tools release
 - with the latest Lexicon and Metathesaurus
- Computer-aided (automatic) process on the sClass tagging
- Add more synonyms from other NLP projects
- Performance tests on NLP applications



Questions



- Lexical Systems Group: <http://umlslex.nlm.nih.gov>
- The SPECIALIST NLP Tools: <http://specialist.nlm.nih.gov>

