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

• 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

<table>
<thead>
<tr>
<th>Terms</th>
<th>Concepts</th>
<th>NLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>cold</td>
<td>Cold Temperature</td>
<td>C0009264</td>
</tr>
<tr>
<td>Cold Temperature</td>
<td>Cold Temperatures</td>
<td>Cold (Temperature)</td>
</tr>
<tr>
<td>Cold Temperatures</td>
<td>Cold (Temperature)</td>
<td>Temperatures, Cold</td>
</tr>
<tr>
<td>Cold (Temperature)</td>
<td>Temperatures, Cold</td>
<td>Cold Temperatures</td>
</tr>
<tr>
<td>Temperatures, Cold</td>
<td>Low temperature</td>
<td>low temperatures</td>
</tr>
<tr>
<td>Cold Temperatures</td>
<td>Low temperature</td>
<td>low temperatures</td>
</tr>
<tr>
<td>Cold Temperatures</td>
<td>low temperatures</td>
<td>...</td>
</tr>
<tr>
<td>Cold Temperature</td>
<td>low temperatures</td>
<td>...</td>
</tr>
<tr>
<td>Cold Temperature</td>
<td>low temperatures</td>
<td>...</td>
</tr>
<tr>
<td>Low temperature</td>
<td>low temperatures</td>
<td>...</td>
</tr>
<tr>
<td>low temperatures</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>cold</td>
<td>Cold Temperature</td>
<td>C0009264</td>
</tr>
<tr>
<td>Cold Temperature</td>
<td>Common Cold</td>
<td>C0009443</td>
</tr>
<tr>
<td>Cold Therapy</td>
<td>C0010412</td>
<td>Cold Sensation</td>
</tr>
<tr>
<td>Cold Sensation</td>
<td>C0234192</td>
<td>...</td>
</tr>
</tbody>
</table>
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
- hodgkin's disease
- Disease;Hodgkins
- Disease, Hodgkin
- ...

Indexed Database

Terms in Corpus

normalize

Index

Indexed Database
Normalized String

• C0019829
• Hodgkin Disease
NLP – Norm (Cont.)

Hodgkin’s Disease

Query → norm → Normed Term → disease hodgkin

Results that matches the normalized query
- C0019829
- Hodgkin Disease

Indexed Database
Normalized String

SQL
Normalized Patterns (in UMLS)

Example: Hodgkin Disease (C0019829)
- There are 139 terms associated with this concept
- Similar normalized patterns

<table>
<thead>
<tr>
<th>Pattern String</th>
<th>No. of Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>hodgkin disease</td>
<td>32</td>
</tr>
<tr>
<td>hodgkin granuloma</td>
<td>22</td>
</tr>
<tr>
<td>hodgkin lymphoma</td>
<td>23</td>
</tr>
<tr>
<td>hodgkin paragranuloma</td>
<td>14</td>
</tr>
<tr>
<td>hodgkin sarcoma</td>
<td>15</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Sub-total</td>
<td>139</td>
</tr>
</tbody>
</table>

Hodgkin Disease
NLP – Query Expansion (Synonym)

Indexed Database Normalized String

Nasal Deformity
- Norm
- deformity nasal

Nose deformity
- Norm
- deformity nose

C0240547
Nose deformity

C0028429:
- nose
- nasal
NLP – Query Expansion (derivation)

Indexed Database Normalized String

perforated ear drum → perforation ear drum

Norm

drum ear perforate → drum ear perforation

Norm

None

C0206504
Tympanic Membrane Perforation
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
entry=E0029526
cat=verb
variants=reg
intran
tran=np
tran=pphr(from,np)
tran=pphr(to,np)
nominalization=generalisation|noun|E0029525
}

spelling variant
part of speech
inflectional variant
chunker
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: generalisation, generalization, generalisable
- prefixD: overgeneralise, 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)

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

<table>
<thead>
<tr>
<th></th>
<th>Recall - Query Expansion</th>
<th>Precision - POS Tagging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>perforated ear drum</td>
<td>saw</td>
</tr>
</tbody>
</table>
| UMLS-CUI             | None                         | C1947903|verb|see
|                      | C0183089|noun|saw (device)            |
| Process              | perforation ear drum         | noun                                     |
| UMLS-CUI             | C0206504                      | C0183089                                 |
| Preferred Term       | Tympanic Membrane Perforation| saw (device)                             |
NLP – Concept Mapping Model

Free Text

Tokenization (Segmentation)
- Documents
- Paragraphs
- Sentences
- Phrases
- Terms (Lexical Lookup)
- Tokens (words)
- ...

Norm Term

Ontology (UMLS) - Indexed Database Normalized Term

CUI
- Yes → WSD (STI)
- No → Ranking

Terms
- Query Expansion (STMT)

(lexical variants, synonyms, co-occurrences, or fruitful variants, etc.)
The SPECIALIST NLP Tools

The SPECIALIST NLP Tools by LSG

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
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
  - \textit{joy|noun|enjoy|verb} \rightarrow \textit{enjoy|verb|joy|noun}

- Transitivity: \(((x = y) \text{ and } (y = z)) \rightarrow (x = z)\)
  - \textit{enjoy|verb|joy|noun}
  - \textit{joy|noun|happy|adj}
  - recursive
  - \textit{enjoy|verb} \rightarrow \textit{joy|noun} \rightarrow \textit{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)
  - C0014901 (esthetics)
  - How about recursive synonym?

```
sensation|noun \leftrightarrow esthetic|adj

esthetic|adj \leftrightarrow 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
  o by resolving issues of using UMLS synonyms
# Broader Issues – Near Synonyms

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

Hinbrain: Metencephalon

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

*the term “brainstem” refers to the medulla oblongata, pons, and the midbrain
## Distinct Issues – Similar but Different

<table>
<thead>
<tr>
<th>CUI</th>
<th>Preferred Term</th>
<th>synonym</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0000741</td>
<td>Abducens nerve structure</td>
<td>abductor digiti minimi</td>
<td>The abductor digiti minimi is a muscle, not a nerve.</td>
</tr>
<tr>
<td>C0003864</td>
<td>Arthritis</td>
<td>arthritide</td>
<td>Per Dorland’s an arthride is “any skin eruption of arthritic or gouty origin.”</td>
</tr>
<tr>
<td>C0005400</td>
<td>Bile duct structure</td>
<td>choledochitis</td>
<td>Choledochitis is a condition of the common bile duct, not structure.</td>
</tr>
<tr>
<td>C0000869</td>
<td>Acacia</td>
<td>locust tree</td>
<td>Though both the acacia &amp; locust tree are members of Leguminosae (pea, bean), they do seem to refer to different trees.</td>
</tr>
<tr>
<td>C0003353</td>
<td>Antigua</td>
<td>Anguilla</td>
<td>The islands of Antigua &amp; Anguilla are both in the West Indies, but are not the same place.</td>
</tr>
</tbody>
</table>
Acacia & Locust tree
Anguilla & Antigua
<table>
<thead>
<tr>
<th>CUI</th>
<th>Preferred Term</th>
<th>synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0001175</td>
<td>Acquired Immunodeficiency Syndrome</td>
<td>sida</td>
</tr>
<tr>
<td>C0001857</td>
<td>AIDS related complex</td>
<td>arc</td>
</tr>
<tr>
<td><strong>C0003023</strong></td>
<td>Angola</td>
<td><strong>ago</strong></td>
</tr>
<tr>
<td>C3714936</td>
<td>Non-Compliant ADaM Datasets Domain</td>
<td>ax</td>
</tr>
</tbody>
</table>
## POS Issues – Meaning Shift

<table>
<thead>
<tr>
<th>CUI</th>
<th>Preferred Term</th>
<th>synonym</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0001774</td>
<td>Agaricales</td>
<td>Mushroom</td>
<td>The verb (to) mushroom does not refer to Agaricales while the noun is a synonym</td>
</tr>
<tr>
<td>C0003459</td>
<td>Anura</td>
<td>frog</td>
<td>The verb (to) frog does not refer to Anura, while the noun is a synonym</td>
</tr>
<tr>
<td>C0003842</td>
<td>Arteries</td>
<td>arterial</td>
<td>The noun arterial refers to roads, not circulatory anatomy, unlike the adjective arterial.</td>
</tr>
<tr>
<td>C0004063</td>
<td>Assault</td>
<td>mug</td>
<td>The verb mug does refer to assault, but the noun mug does not.</td>
</tr>
</tbody>
</table>
Recursive Issues – Multiple Concepts

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

<table>
<thead>
<tr>
<th>CUIs</th>
<th>Synonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0009443 common cold</td>
<td>• cold</td>
</tr>
<tr>
<td></td>
<td>• coryza</td>
</tr>
<tr>
<td></td>
<td>• acute coryza</td>
</tr>
<tr>
<td></td>
<td>• common cold</td>
</tr>
<tr>
<td>C0009264 cold temperature</td>
<td>• low temperature</td>
</tr>
<tr>
<td></td>
<td>• low-temperature</td>
</tr>
<tr>
<td></td>
<td>• lowtemperature</td>
</tr>
<tr>
<td>C0234192 cold sensation</td>
<td>• psychoesthesia</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Synonyms</th>
<th>cold temperature</th>
<th>cold therapy</th>
<th>common cold</th>
<th>cold sensation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1-G substitution</strong></td>
<td>cold temperature</td>
<td>cold therapy</td>
<td>common cold</td>
<td>cold sensation</td>
</tr>
<tr>
<td><strong>2-G substitution</strong></td>
<td>cold temperature temperature cold therapy temperature common cold temperature cold sensation temperature</td>
<td>cold temperature therapy cold therapy therapy common cold therapy common cold therapy cold sensation therapy</td>
<td>common cold temperature common cold therapy common cold therapy common cold sensation</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LexSynonymyms – 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

<table>
<thead>
<tr>
<th>#SYNONYM_CLASS</th>
<th>C0003842</th>
<th>Arteries</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>E0010481</td>
<td>arteria</td>
</tr>
<tr>
<td>128</td>
<td>E0010531</td>
<td>artery</td>
</tr>
<tr>
<td>128</td>
<td>E0694191</td>
<td>arterial</td>
</tr>
<tr>
<td>1</td>
<td>E0010482</td>
<td>arterial</td>
</tr>
</tbody>
</table>

#SYNONYM_CLASS | C0004063 | Assault |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>E0041250</td>
<td>mug</td>
</tr>
<tr>
<td>128</td>
<td>E0010822</td>
<td>assault</td>
</tr>
<tr>
<td>128</td>
<td>E0041249</td>
<td>mug</td>
</tr>
</tbody>
</table>

...
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=death
entry=E0020877
  cat=adj
  variants=inv
  variants=reg
  position=attrib(1)
  position=attrib(3)
  position=pred
  stative
  nominalization=deadness|noun|E0020885
}

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

{base=dead
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
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
dead|1|die|1024|C0011065
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
dead|1|die|1024|C0011065
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
dead|1|die|1024|C0011065
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
dead|1|die|1024|C0011065
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
dead|1|die|1024|C0011065
dead|1|death|128|C0011065
dead|1|deceased|1|C0011065
LexSynonym Generation

- Retrieve synonym candidates (sClasses)
- Tag sClasses
- Generate sPairs (CUI)
- Generate sPairs from nominalizations (EUI)
- Generate sPairs from Lexical Tools, 2016 (LVG)
- Combine and convert to sPairs
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

CUI
CUI: C0011065
PT: Cessation of life
- die
- dead
- death
- deadness
- deceased

NLP
CUI: C0231800
PT: Expiration
- die
- expire
- deceased

EUI
X

- terminate
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 Metathesaurus, 2016 Lexicon

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Tagged</th>
<th>Completion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sClass</td>
<td>22,779</td>
<td>7,686</td>
<td>33.74%</td>
</tr>
<tr>
<td>Synonyms</td>
<td>80,913</td>
<td>29,990</td>
<td>37.06%</td>
</tr>
</tbody>
</table>

- Synonym stats:

<table>
<thead>
<tr>
<th>Year</th>
<th>CUI</th>
<th>EUI</th>
<th>NLP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5,198 (100%)</td>
<td>5,198</td>
</tr>
<tr>
<td>2017</td>
<td>118,468 (62%)</td>
<td>67,584 (35%)</td>
<td>4,792 (3%)</td>
<td>190,844</td>
</tr>
</tbody>
</table>

36.71 growth
# Summary

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

- Complete tagging on all sClasses

- 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

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Tagged</th>
<th>Completion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sClass</td>
<td>22,779</td>
<td>7,686</td>
<td>33.74%</td>
</tr>
<tr>
<td>Synonyms</td>
<td>80,913</td>
<td>29,990</td>
<td>37.06%</td>
</tr>
</tbody>
</table>
Questions