

# **Lexical Tools Sub-Term Mapping Tools (STMT)**

Dr. Chris J. Lu

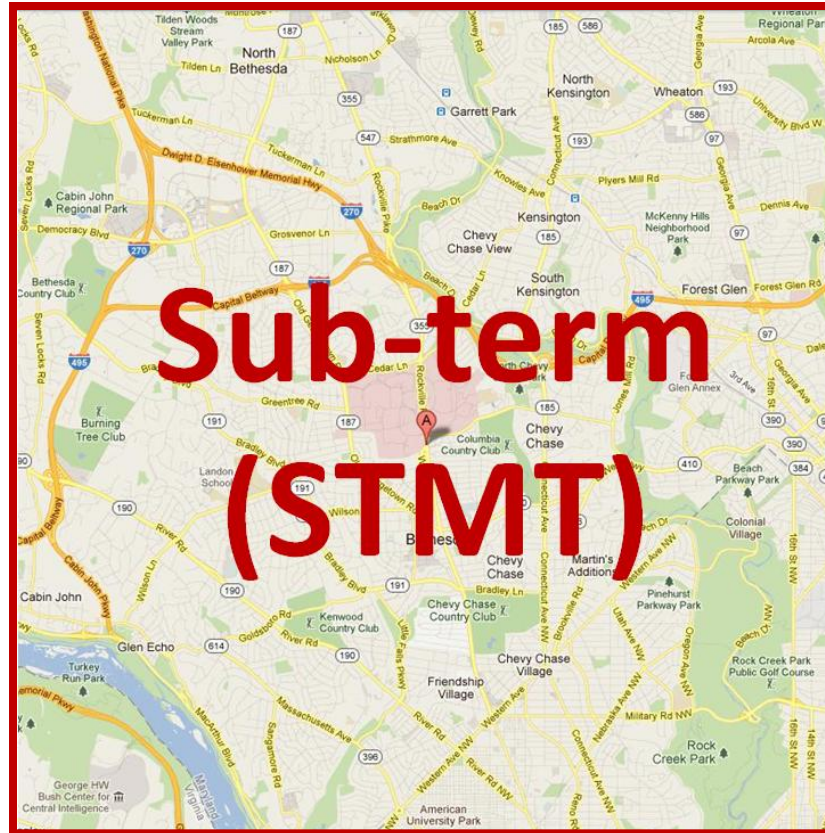
[The Lexical Systems Group](#)  
[NLM](#). [LHNCBC](#). [CGSB](#)

April, 2012

# Table of Contents

- Introduction
  - MetaMap
  - UMLS-Core
- STMT
  - Corpus & Normalization
  - Sub-Term
  - Sub-Term Patterns
  - Synonyms Substitutions
- Current Plan
- Questions

# STMT - Introduction



# Sub-term

- A sub-term is a term that is a subset of another term
- A term is composed of words
- A word is delimited by space(s) or tab(s)

# Sub-term: Example

- Term: otitis externa, chronic infectious
- Corpus: The SPECIALIST Lexicon
- Sub-terms:
  - otitis (E0044452)
  - otitis externa (0044453)
  - chronic (E0219527|E0016869)
  - infectious (E0044452)

# Application – MetaMap\*

- Longest Prefix:
  - Find the longest prefix sub-term on an input term that is a lexical item (LexItem) in the SPECIALIST Lexicon
  - Ignore case and punctuation

\* A.R. Aronson and F.M. Lang, “An Overview of MetaMap: historical perspective and recent advances”, JAMIA, 2010, Vol. 17, p.229-236

# Lexical Mapping

- Algorithm:
  1. Normalization (LexItem Norm, -f:g:rs:o:l)
    - One to one
  2. Lexicon Mapping (LexAccess)
    - Table lookup (key - values)

- Example:

University of California, Los Angeles

⇒ university of california los angeles (E0063261)

⇒

Original Term from Lexicon	Normalized Term (key)	EUI
...	...	...
University of California, Los Angeles	university of california los angeles	E0063261
University of California Los Angeles	university of california los angeles	E0063261
University of California-Los Angeles	university of california los angeles	E0063261
...	...	...

# Lexical Mapping (cont.)

- University of California, Los Angeles
- University of California Los Angeles
- University of California-Los Angeles
- university of california los angeles
- UNIVERSITY OF CALIFORNIA LOS ANGELES
- ...

Normalization (LexItem Norm: -f:g:rs:o:l)

university of california los angeles

LexAccess (lvg -f:E)

Original Term from Lexicon	Normalized Term (key)	EUI
...	...	...
University of California, Los Angeles	university of california los angeles	E0063261
University of California Los Angeles	university of california los angeles	E0063261
University of California-Los Angeles	university of california los angeles	E0063261
...	...	...

E0063261



# Longest Prefix

- University of California, Los Angeles is in USA ...
- University of California Los Angeles is in USA ...
- University of California-Los Angeles is in USA ...
- university of california los angeles is in USA ...
- UNIVERSITY OF CALIFORNIA LOS ANGELES is in USA ...
- ...

Normalization (LexItem Norm: -f:g:rs:o:l)

university of california los angeles is in usa ...

Lexical Mapping on each prefix sub-term

# Longest Prefix (cont.)

- University of California, Los Angeles is in USA ...
  - Normalize (LexItem Norm)  
=> university of california los angeles is in usa ...
  - Word by word
    1. university (E0063257)
    2. university of (**none**)
    3. university of california (E0702384)
    4. university of california, los (**none**)
    5. university of california, los angles (E0063261)
    6. university of california, los angles is (**none**)
    7. ...

# Longest Prefix (cont.)

- university of california los angeles is in usa ...
  - I. Normalized
  - II. Word by word
    1. university (E0063257)
    2. university of (none)
    3. university of california (E0702384)
    4. university of california, los (none)
    5. university of california, los angles (E0063261)
    6. university of california, los angles is (none)
    7. ...
- Issues:
  - Performance
  - Limited to 10 prefixes

# Application – UMLS-CORE\*

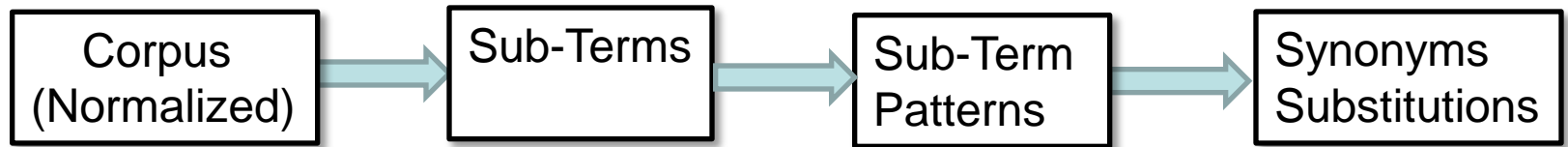
- Query expansion: to find concepts in 3 steps:
  - 1) Normalized lexical matches
    - MRXNS\_ENG.RRF: normalized string to CUI
    - Lexical tools – Norm (-f:q0:g:rs:o:t:l:B:Ct:q7:q8:w)
  - 2) One synonymous word or phrase substitution
  - 3) Two synonymous word or phrase substitutions
    - acronyms: AB|ANTIBODY
    - British: aeroplane|airplane
    - Greco-latin: cat|feline
    - ECRI (devices): Airway|Tube
    - Lvg: zoic|animal life
    - ...

\* K.W. Fung, C. McDonald, S. Srinivasan, "The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions", JAMIA, 2010, Vol. 17, p.675-680

# UMLS-Core: Example

- OTITIS EXTERNA, CHRONIC INFECTIOUS
  - 1) Normalized lexical matches
    - chronic externa infectious otitis
    - chronic externon infectious otitis
    - chronic externum infectious otitis⇒ No CUI found
  - 2) One synonymous word or phrase substitution  
⇒ No CUI found
  - 3) Two synonymous word or phrase substitutions  
⇒ ...

# STMT



# STMT - MetaMap

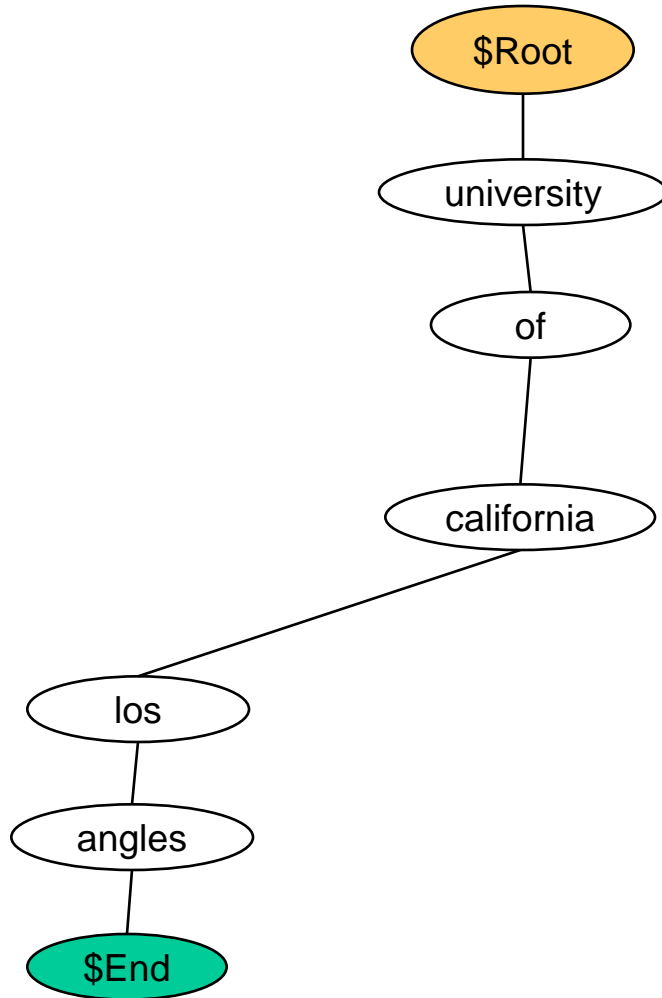


- Corpus (Lexicon) – LexItem Norm

Original Term from Lexicon	Normalized Term (key)	EUI
...	...	...
University of California, Los Angeles	university of california los angeles	E0063261
University of California Los Angeles	university of california los angeles	E0063261
University of California-Los Angeles	university of california los angeles	E0063261
University of California, San Francisco	university of california san francisco	E0063262
University of California SanFrancisco	university of california sanfrancisco	E0063262
University of California-SanFrancisco	university of california sanfrancisco	E0063262
University of California	university of california	E0702384
University of California, Davis	university of california davis	E0702385
University of California Davis	university of california davis	E0702385
University of California-Davis	university of california davis	E0702385
University of California, San Diego	university of california san diego	E0702386
...	...	...

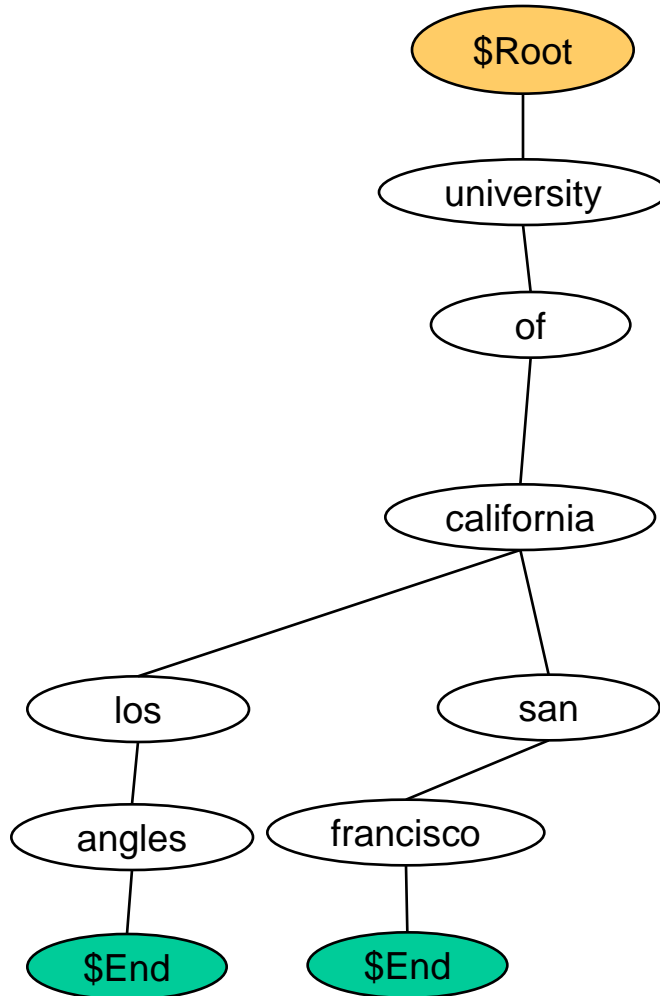
- Load the corpus to a tree structure with each term as a branch in the tree and each word in the term as a node in the branch

# STMT

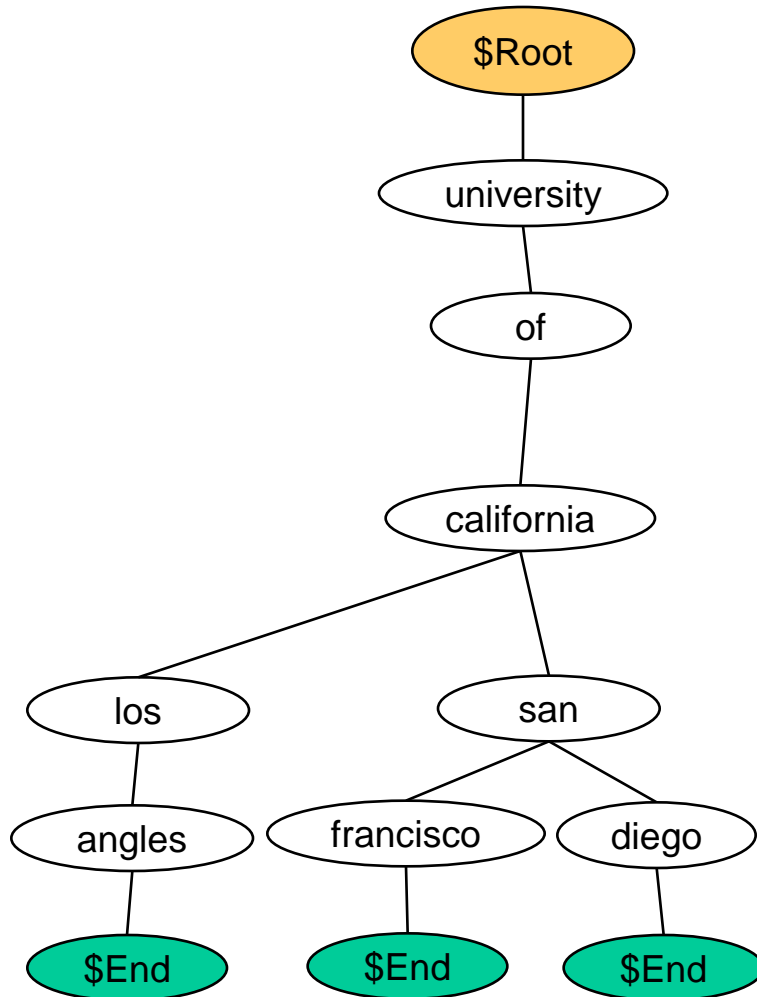




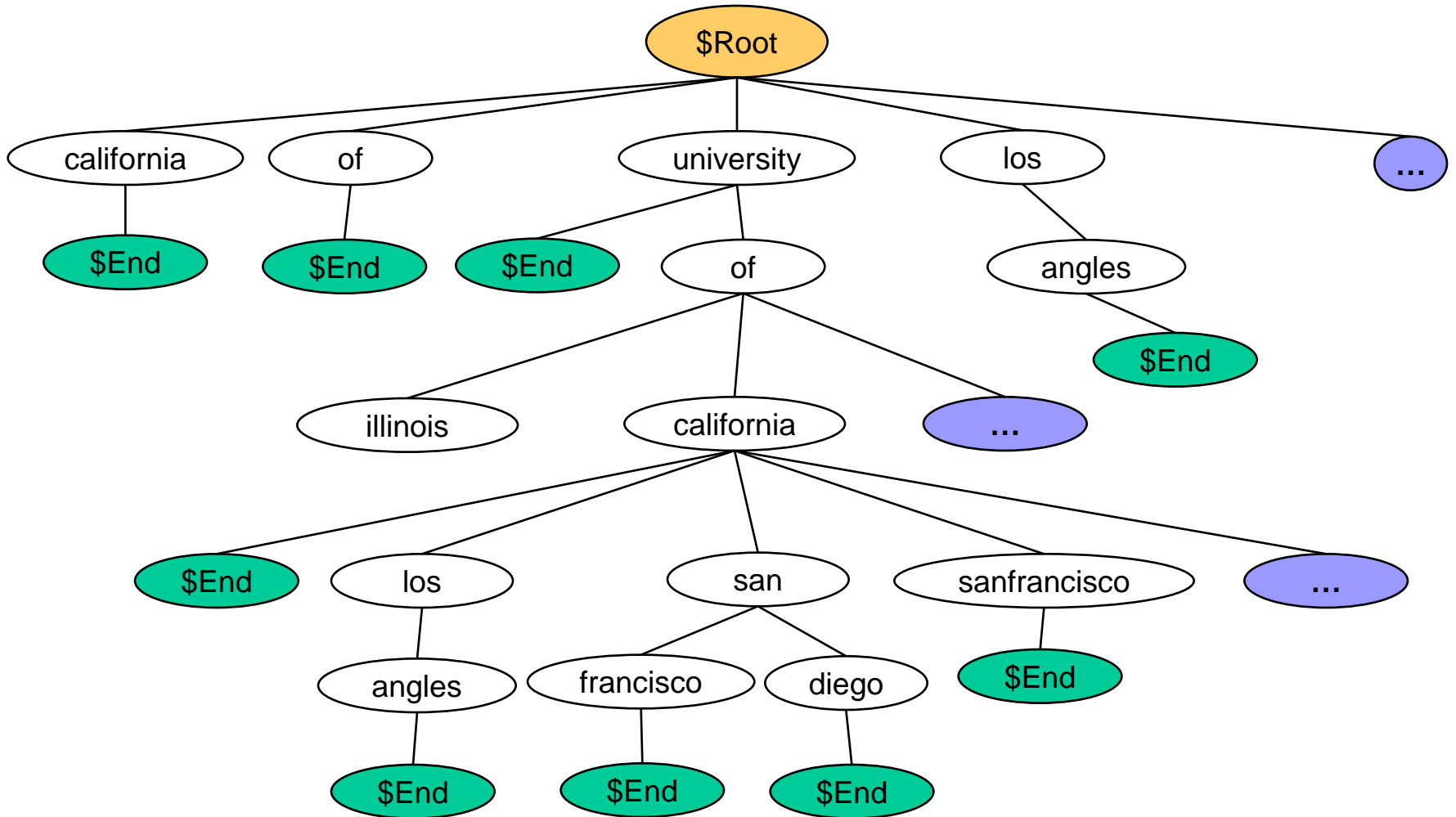
# STMT



# STMT



# STMT



# STMT – Longest Prefix

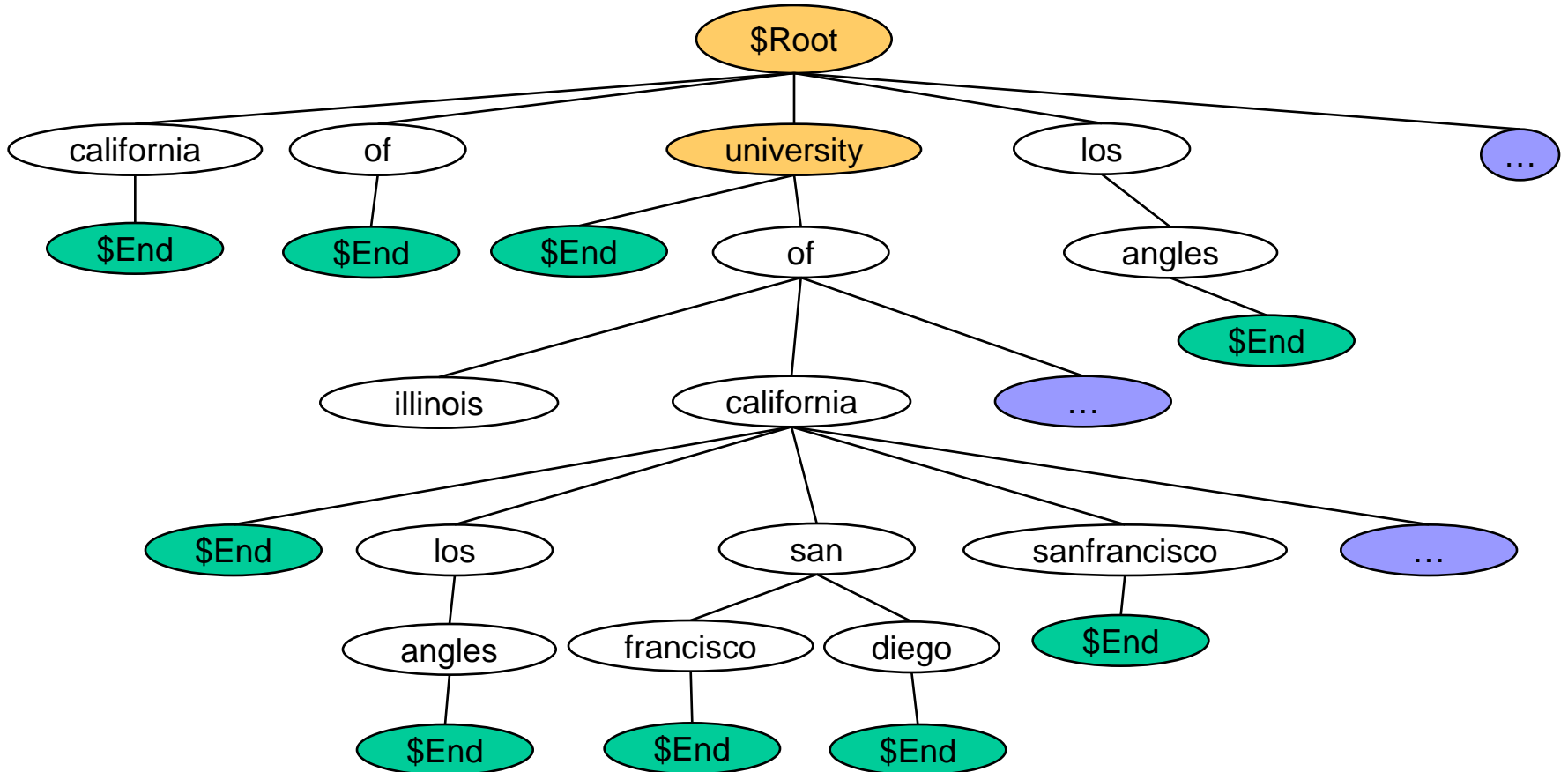


university of california los angeles is in usa ...

# STMT – Longest Prefix



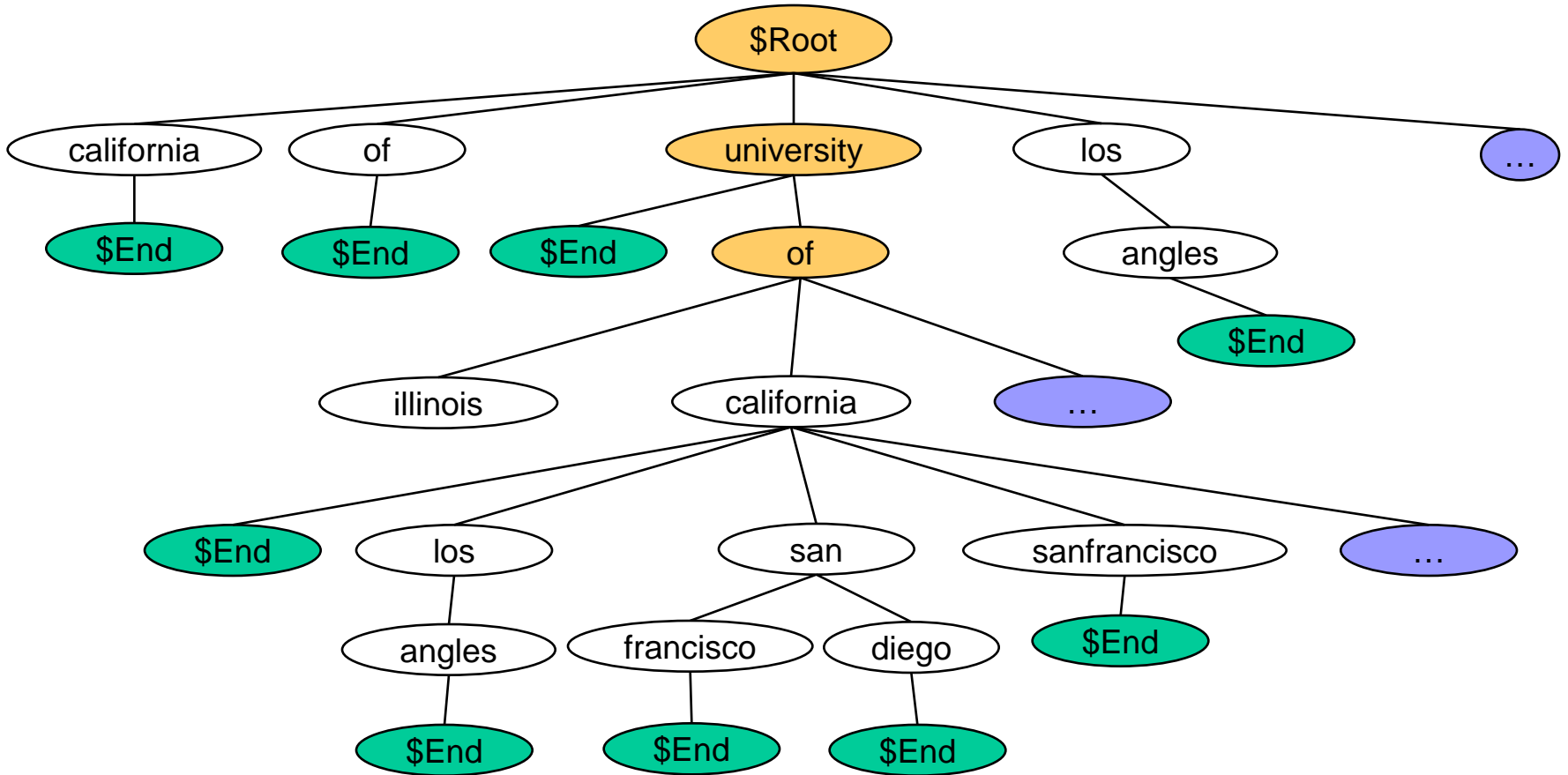
university of california los angeles is in usa ...



# STMT – Longest Prefix



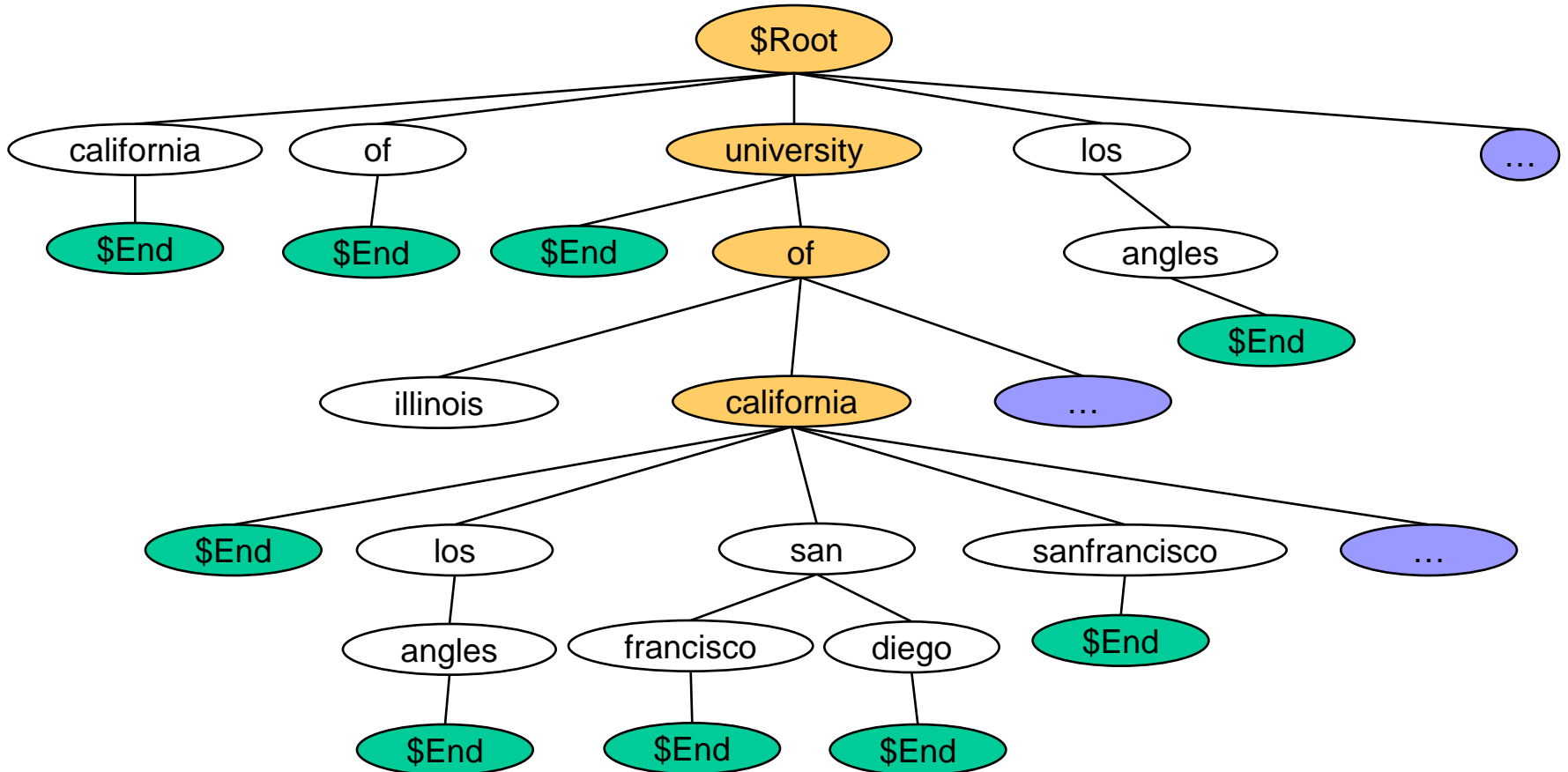
university of california los angeles is in usa ...



# STMT – Longest Prefix



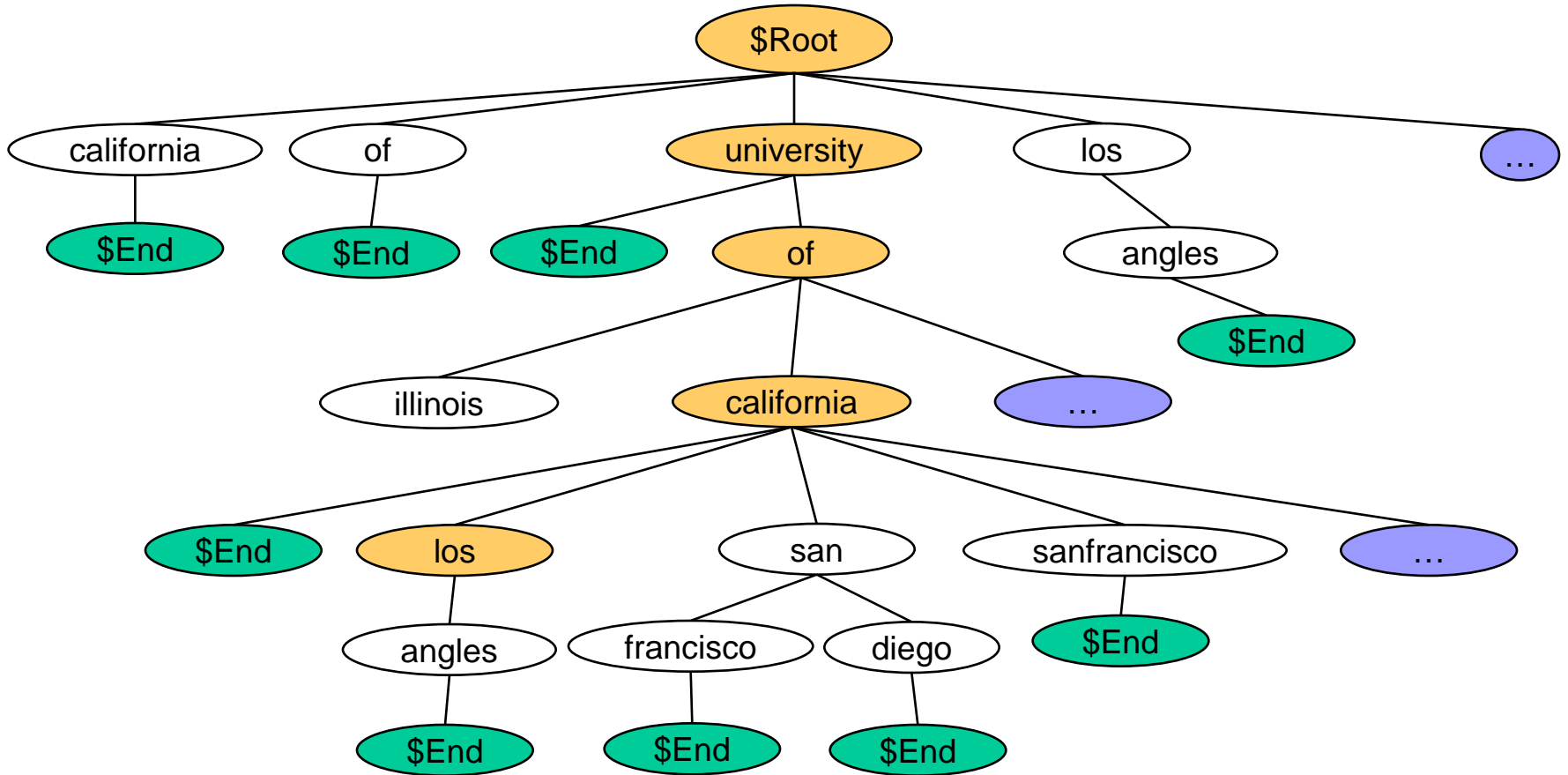
university of california los angeles is in usa ...



# STMT – Longest Prefix



university of california los angeles is in usa ...

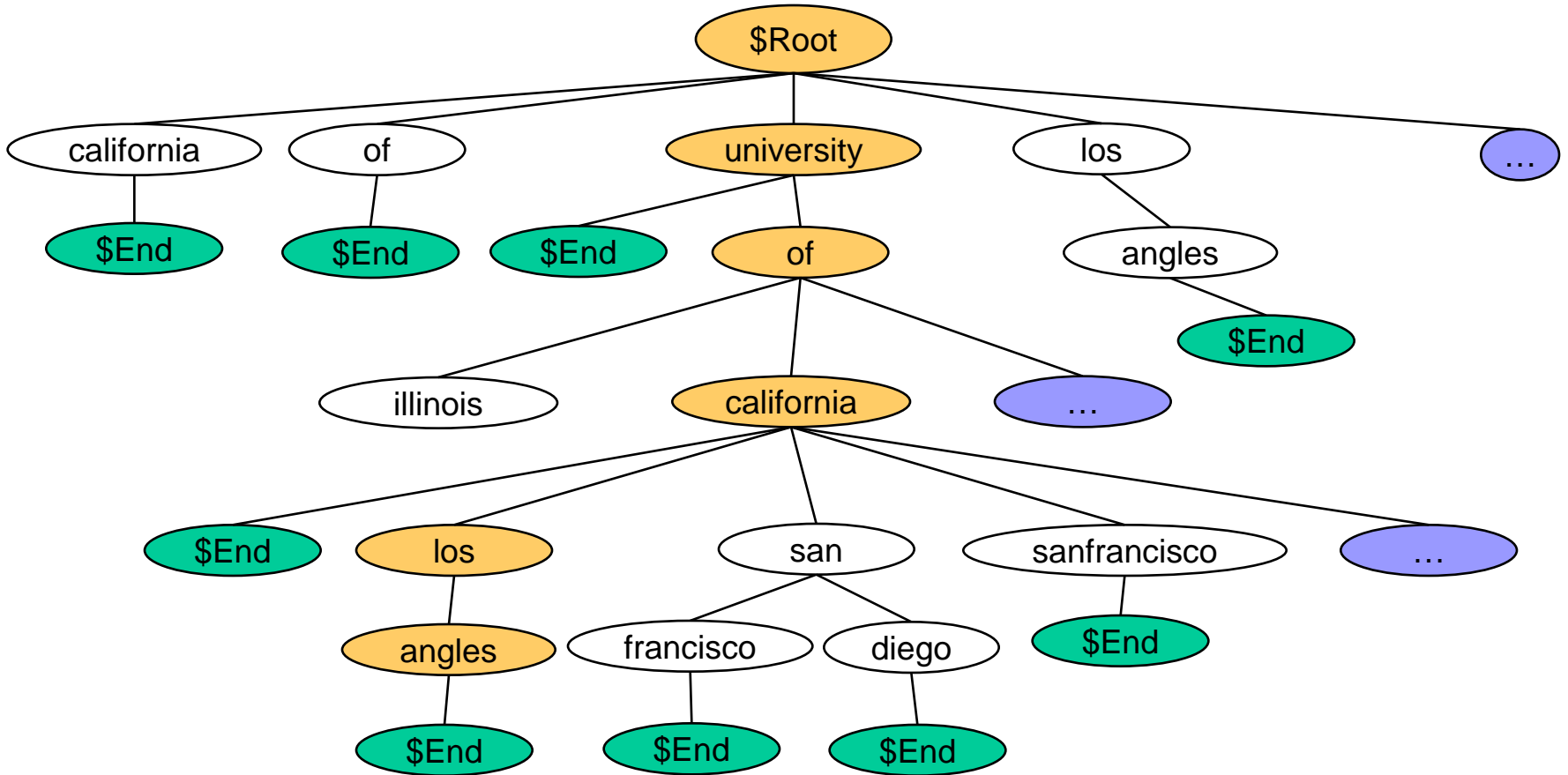




# STMT – Longest Prefix



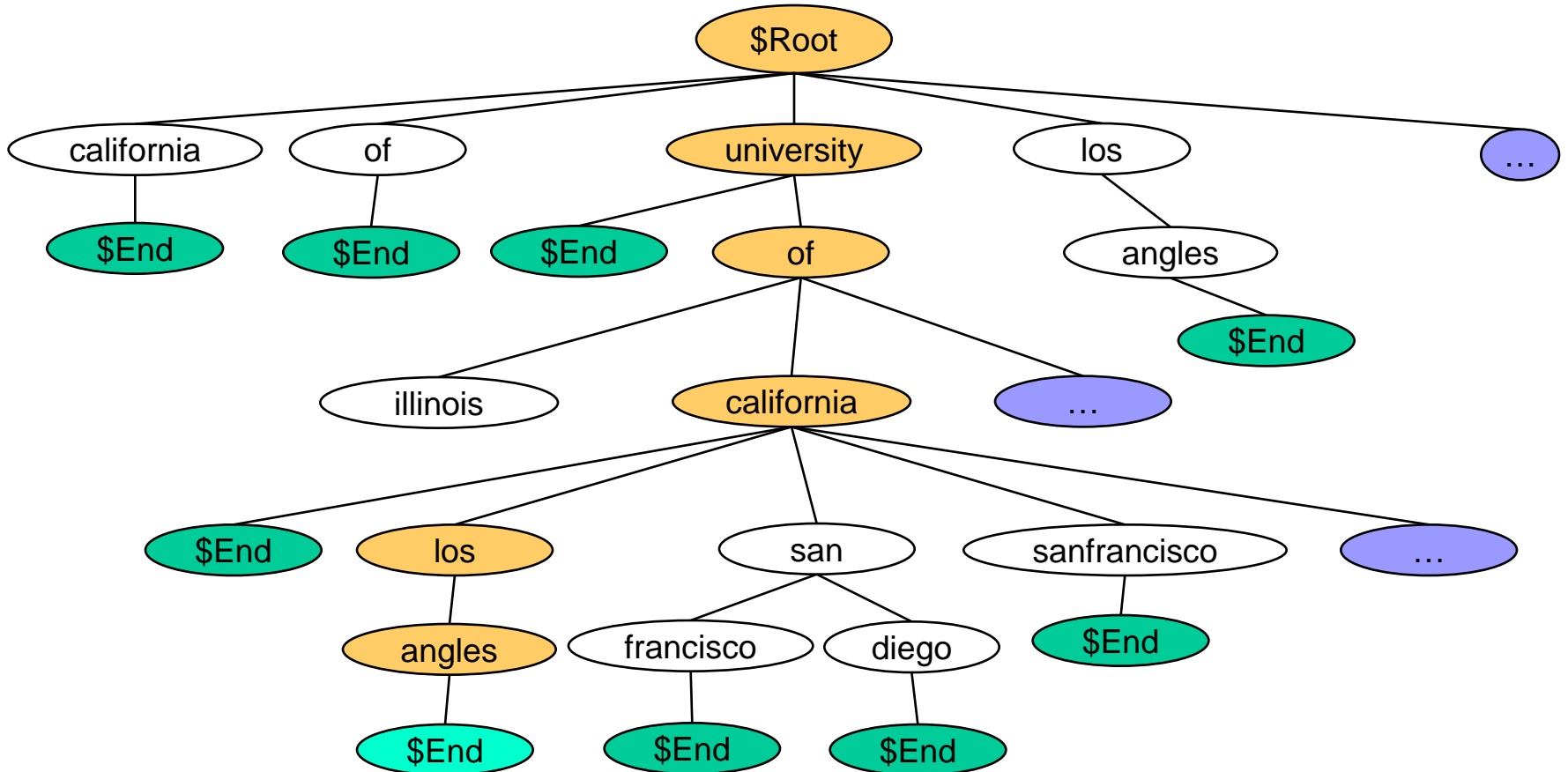
university of california los angeles is in usa ...



# STMT – Longest Prefix



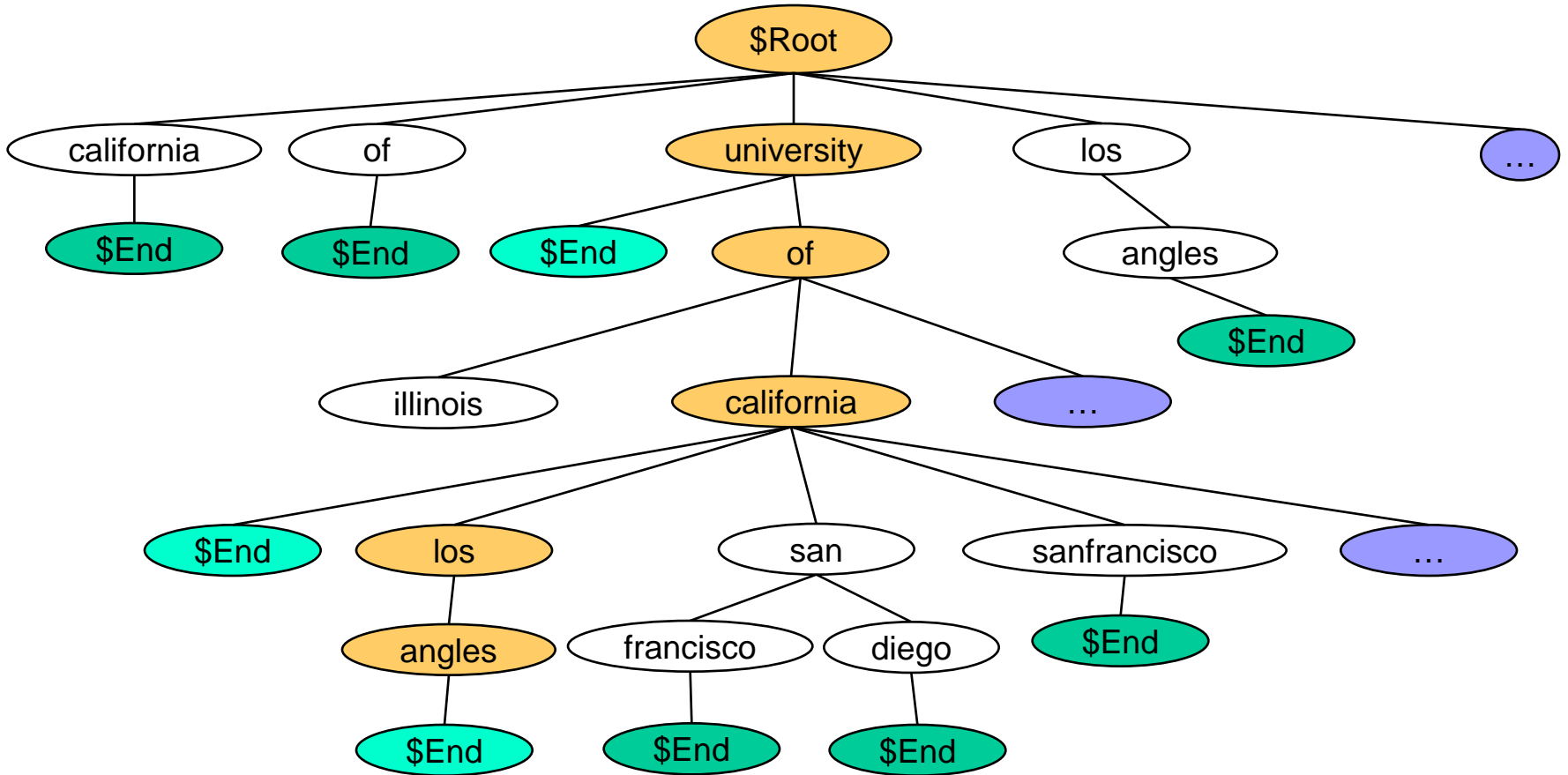
university of california los angeles is in usa ...



# STMT – Sub-terms



university of california los angeles is in usa ...



# STMT – UMLS-Core



- OTITIS EXTERNA, CHRONIC INFECTIOUS
  - 1) Normalized lexical matches
    - chronic externa infectious otitis => no CUI found
    - chronic externon infectious otitis => no CUI found
    - chronic externum infectious otitis => no CUI found
  - 2) One sub-term (synonymous word or phrase) substitution
  - 3) Two sub-term substitutions

# STMT – UMLS-Core



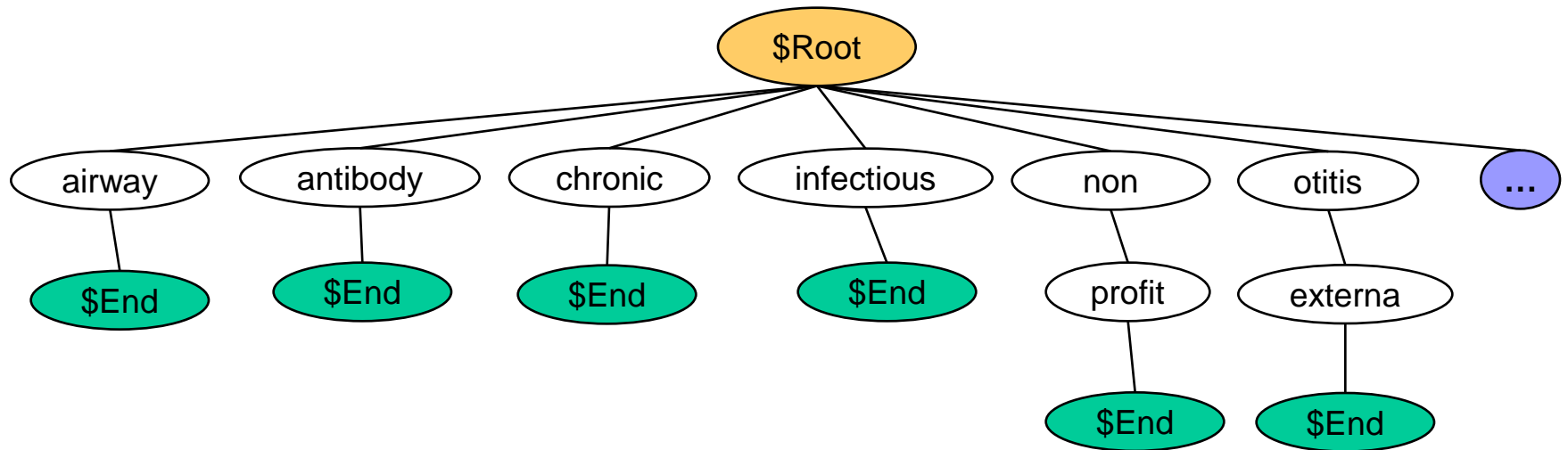
- Corpus (Synonyms files):
  - Acronyms, British, Greco-latin, ECRI, Lvg, etc.
- Synonym Norm (-f: g:rs:Ct:o:l)
  - Abstract away from punctuation, cases, spelling variants, inflectional variants
  - One to many
  - Remove from synonym list if the synonym pair are the same after norm (side-to-side | side to side)
- Load the normalized synonyms to corpus tree

# STMT – UMLS-Core



Original Term from Synonyms	Normalized Term (key)	synonyms
...	...	...
chronic	chronic	chron
chronic	chronic	long-term
chronic	chronic	persistent
chronic	chronic	recurrent periodic
chronic	chronic	relapsing
infectious	infectious	communicable
infectious	infectious	contagious
infectious	infectious	infection
otitis externa	otitis externa	auditory
otitis externa	otitis externa	auditory canal
otitis externa	otitis externa	aural
otitis externa	otitis externa	ear
ANTIBODY	antibody	AB
Airway	airway	Tube
non-profit	non profit	not for profit
...	...	...

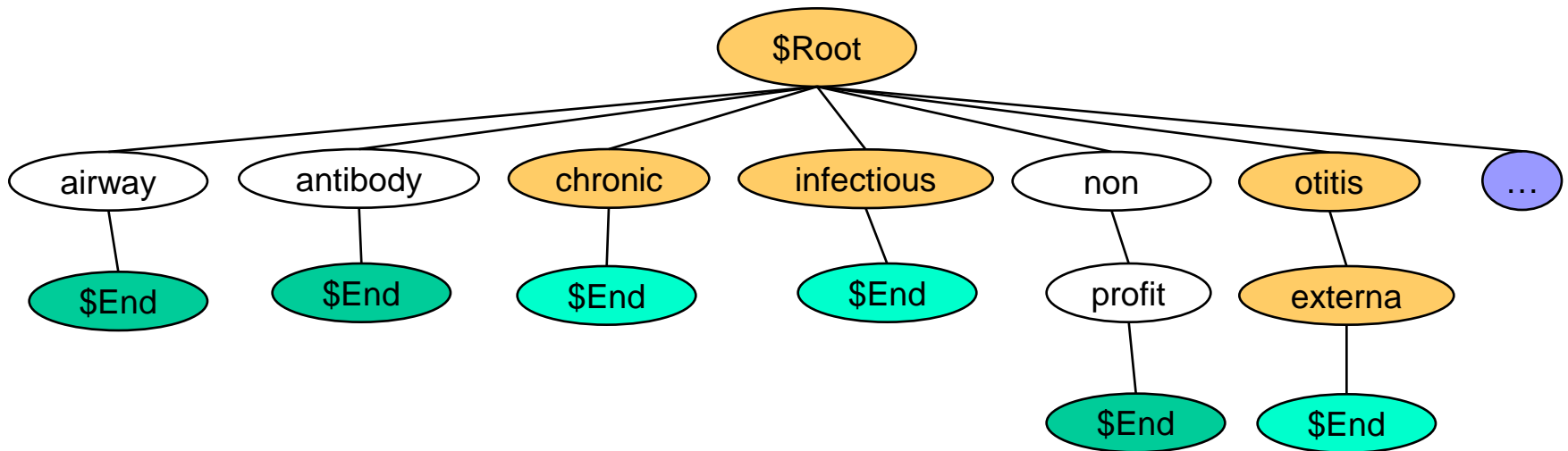
# STMT – UMLS-Core



# STMT – UMLS-Core



- OTITIS EXTERNA, CHRONIC INFECTIOUS  
=> otitis externa chronic infectious
- Three sub-terms found
  - otitis externa|0|2
  - chronic|2|3
  - infectious|3|4





# STMT – UMLS-Core



- OTITIS EXTERNA, CHRONIC INFECTIOUS
- => otitis externa chronic infectious
- Sub-term patterns with one substitution
  - otitis externa chronic infectious
  - otitis externa chronic infectious
  - otitis externa chronic infectious
- Sub-term patterns with two substitutions
  - otitis externa chronic infectious
  - otitis externa chronic infectious
  - otitis externa chronic infectious
- It could be really complicated (multiple overlaps)

# STMT – UMLS-Core



- otitis externa chronic infectious
- Synonyms:
  - otitis externa (4): auditory | auditory canal | aural | ear
  - chronic (5): chron | long-term | persistent | relapsing | recurrent periodic
  - infectious (3): communicable | contagious | infection
- Sub-term patterns with one substitution (12)
  - otitis externa chronic infectious (4)
  - otitis externa chronic infectious (5)
  - otitis externa chronic infectious (3)
- Sub-term patterns with two substitutions (47)
  - otitis externa chronic infectious (20)
  - otitis externa chronic infectious (12)
  - otitis externa chronic infectious (15)

# STMT – UMLS-Core



- otitis externa chronic infectious
- Sub-term patterns with one substitution (12)
  - otitis externa chronic infectious (4)
  - otitis externa chronic infectious (5)
  - otitis externa chronic infectious (3)

=> No CUI Found
- Sub-term patterns with two substitutions (47)
  - otitis externa chronic infectious (20)
  - otitis externa chronic infectious (12)
  - otitis externa chronic infectious (15)

=> ear chronic infection | C0743359 | EAR INFECTION CHRONIC

# STMT Plan

- First release: STMT, 2013
- Provide Java APIs, command line tools
  - Generic tool
    - STMT (Sub-Term Mapping Tool)
  - Preloaded corpus
    - LSF (LexItem Sub-Term Finder)
    - SMT (Synonym Mapping Tool)
    - ...
- Web site with full documentations

# Current Status/Future Work

- Major software components are completed
- Future work:
  - Tests
  - Web site - documents
  - Deployable package
  - Persistent corpus

# References

- Lu, Chris J. and Browne, Allen C., "[Development of Sub-Term Mapping Tools](#)", Submitted for publication in Proceeding of AMIA 2012 Annual Symposium, Nov. 3-7, 2012, Chicago, IL
- <http://umlslex.nlm.nih.gov/stmt> (TBD)

# Questions



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