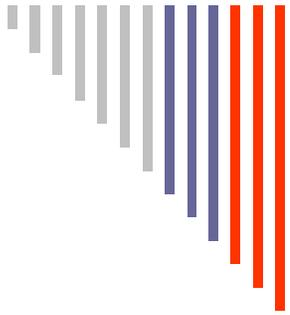


Exploring WordNet as a Consumer Health Vocabulary Source:

**A comparison of arthritis term
coverage in WordNet and the UMLS**

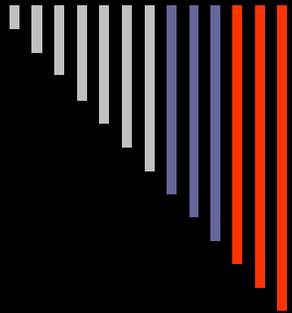


Catherine K. Craven, MLS

NLM Biomedical and Health
Informatics Training Grant Fellow

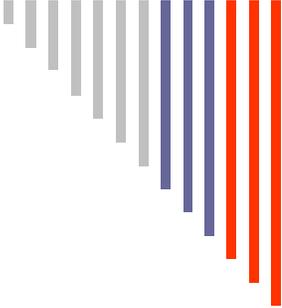
School of Medicine,
Department of Health Management and Informatics,
University of Missouri
324 Clark Hall
Columbia, MO 65211
cravenc@health.missouri.edu

Advisor: Timothy B. Patrick, Ph.D.



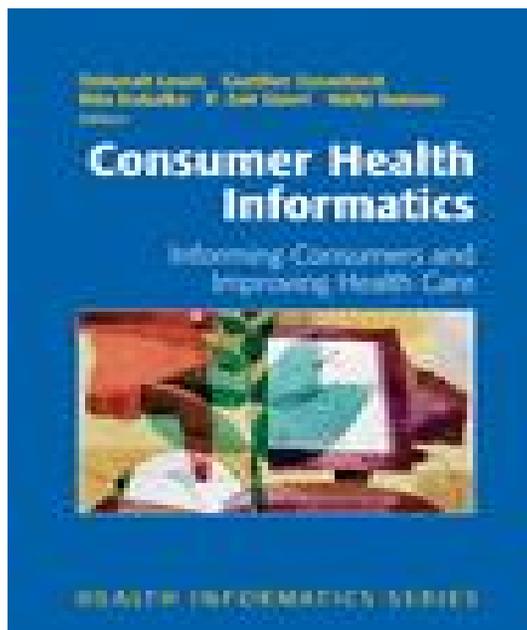
NLM Summer Informatics Rotation Mentors:

- Olivier Bodenreider, M.D.
Ph.D.
 - Tony Tse, Ph.D.
-



Presentation outline

- ❑ Consumer health vocabularies -- why are they important for informatics?
- ❑ What is WordNet and why map to it to explore it for consumer arthritis terms?
- ❑ Why map these consumer arthritis terms to the UMLS Metathesaurus?
- ❑ How we did it;
- ❑ Results and discussion;
- ❑ Future project directions.



Consumer Health Vocabularies

“The ‘consumer vocabulary problem’ {is} a fundamental issue in health information provision...the mismatch between terms used by healthcare professionals and those used by consumers who receive their services.”

“Eight in ten internet users have looked online for health information....That translates into about 95 millions American adults (18+) who use the internet to find health information.”

PEW / INTERNET
PEW INTERNET & AMERICAN LIFE PROJECT

PMVs and CMVs

Real medical
professional
discourse

Different groups:
different “talk”

The way real
non-experts
talk

Varies:
trained professionals,
medical texts,
dictionaries, etc.

Sources of
vocabulary terms

Log queries, e-mail,
health chat forums

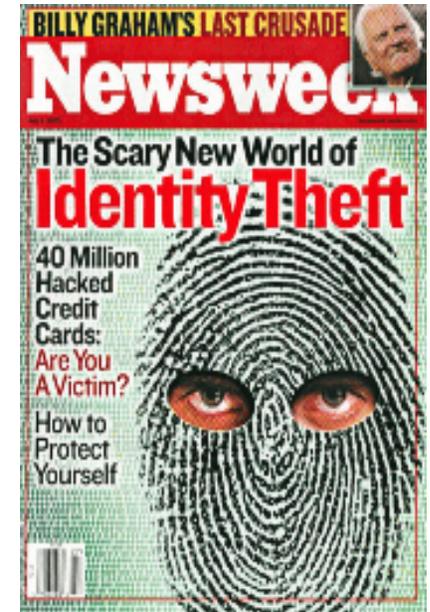
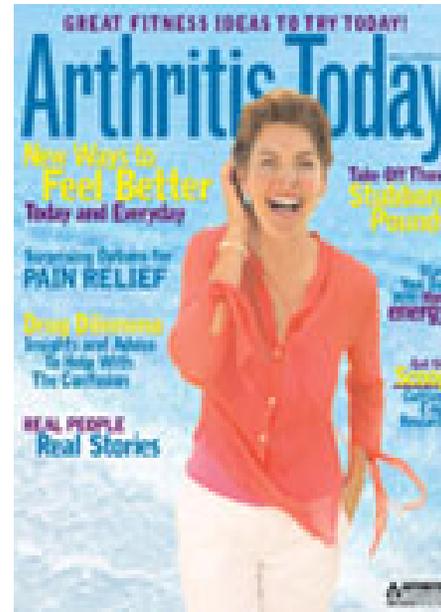
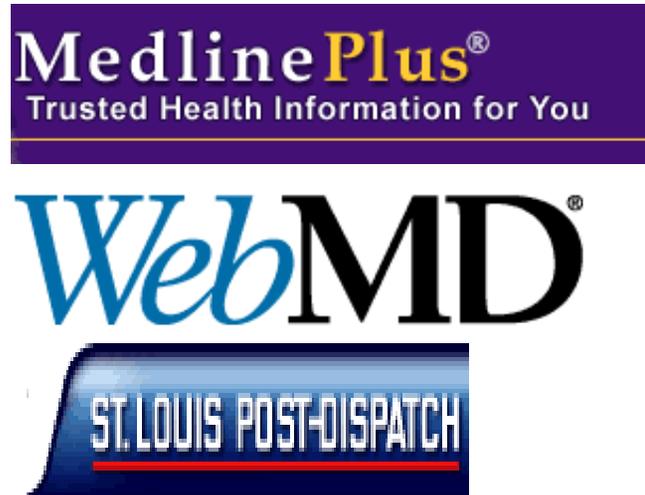
e.g. MeSH,
Snomed CT,
ICD-9 CM

Resulting
vocabularies
(...*after* just a
little bit of work!)

Consumer
Health
Vocabularies

MMVs

Example MMV sources:



Possible MMV source genres:

Informational, Persuasive, Commercial, Human Interest

Possible MMV source purposes:

Improving understand; influencing behavior/thoughts; selling; eliciting interest

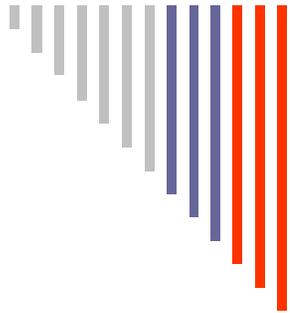


Why the medical domain of arthritis?



says:

- “Arthritis is the leading cause of disability in the United States, limiting the activities of more than 16 million adults.”
- “The total cost of AORC in the United States in 1997 was \$86.2 billion (including \$51.1 billion in direct costs and \$35.1 billion in indirect costs), approximately 1% of the U.S. gross domestic product.”



Identifying a consumer health corpus on arthritis

- One WebMD article for manual “extraction” of terms
- ~ 31,000 word document corpus comprising 17 arthritis articles from various consumer health sources for automatic term extraction

WordNet - Download - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://wordnet.princeton.edu/obtain> Go Links

WordNet

a lexical database for the English language

Cognitive Science Laboratory Princeton University 221 Nassau St. Princeton, NJ 08542

About WordNet	<p>The most recent Windows version of WordNet is 2.1, released in March 2005. Version 2.1 for Unix/Linux/Solaris/etc. was released in July 2005, as was the sense-mapping package. The Prolog version will be released soon.</p> <p>WordNet binaries and source are available for Windows and Unix-like systems (Irix, Solaris, and Linux binaries). There is also a Prolog package.</p> <h3>WordNet 2.1 for Windows</h3> <p>WordNet browser, command-line tool, and database files with InstallShield self-extracting installer:</p> <p>Download via HTTP: WordNet-2.1.exe or FTP: WordNet-2.1.exe</p> <h3>WordNet 2.1 for UNIX-like systems (including: Linux, Mac OS X, Solaris)</h3> <p>Before you download: The WordNet 2.1 README file contains additional information about the release. You can read about the changes from version 2.0.</p> <p>Source code and binaries:</p> <p>Download via HTTP: WordNet-2.1.tar.gz or FTP: WordNet-2.1.tar.gz</p> <h3>Prolog version of WordNet 2.0</h3> <p>ANSI Prolog version of the WordNet database.</p>
Use WordNet online	
Download	
Frequently Asked Questions	
Related projects	
WordNet documentation	
WordNet statistics	
Publications	
License and commercial use	
Contact, Report bugs	

start | Inbox - Micro... | 2 Windows ... | Microsoft Po... | Microsoft Exc... | 2 Microsoft ... | 2 Internet ... | 2:40 PM

Search Word: Searches for bones: Senses:

The noun bones has 1 sense (no senses from tagged texts)

1. {02841411} <noun.artifact> **bones**#1, castanets#1, clappers#1, finger cymbals#1, maraca#1 -- (a percussion instrument consisting of a pair of hollow pieces of wood or bone (usually held between the thumb and fingers) that are made to click together (as by Spanish dancers) in rhythm with the dance)

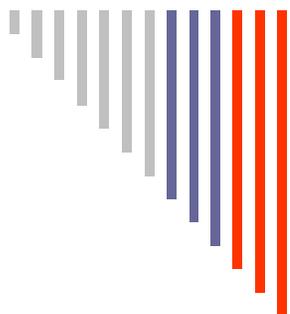
The noun bone has 3 senses (first 2 from tagged texts)

1. (10) {05206933} <noun.body> **bone**#1, os#5 -- (rigid connective tissue that makes up the skeleton of vertebrates)
2. (3) {14563523} <noun.substance> **bone**#2, osseous tissue#1 -- (the porous calcified substance from which bones are made)
3. {04904454} <noun.attribute> **bone**#3, ivory#2, pearl#2, off-white#1 -- (a shade of white the color of bleached bones)

The verb bone has 2 senses (no senses from tagged texts)

1. {00598187} <verb.cognition> cram#3, grind away#1, drum#3, bone up#1, swot#1, get up#3, mug up#1, swot up#1, **bone**#1 -- (study intensively, as before an exam; "I had to bone up on my Latin verbs before the final exam")
2. {00194444} <verb.change> **bone**#2, debone#1 -- (remove the bones from; "bone the turkey before roasting it")

Overview of bones



WordNet 2.1 statistics

Part of Speech Pairs	Unique Stings	Synsets	Total Word-Sense
Noun	117,097	81,426	145,104
Verb	11,488	13,650	24,890
Adjective	22,141	18,877	31,302
Adverb	4,601	3,644	5,720
Totals	155,327	117,597	207,016

Search Word: bones

Searches for bones: Noun Verb

Senses:

The noun bones has 1 sense (no senses from tagged texts)

1. (02841411) <noun> bone (usually he

The noun bone has 3 senses (no senses from tagged texts)

1. (10) (05206933) <noun> bone
2. (3) (14563523) <noun> bone
3. (04904454) <noun> bone

- Synonyms, ordered by estimated frequency
- Coordinate Terms
- Hypernyms (bones is a kind of...)
- Hyponyms (...is a kind of bones), brief
- Hyponyms (...is a kind of bones), full
- Holonyms (bones is a part of...), regular
- Meronyms (parts of bones), regular
- Meronyms (parts of bones), inherited
- Derivationally related forms
- Domain
- Familiarity

#1, finger cymbals#1, maraca#1 -- (a percussion instrument consisting of a pair of hollow pieces of wood or made to click together (as by Spanish dancers) in rhythm with the dance)

ective tissue that makes up the skeleton of vertebrates)
1 -- (the porous calcified substance from which bones are made)
-white#1 -- (a shade of white the color of bleached bones)

The verb bone has 2 senses (no senses from tagged texts)

- 1. (00598187) <verb.cognition> cram#3, grind away#1, drum#3, bone up#1, swot#1, get up#3, mug up#1, swot up#1, **bone**#1 -- (study intensively, as before an exam; "I had to bone up on my Latin verbs before the final exam")
- 2. (00194444) <verb.change> **bone**#2, debone#1 -- (remove the bones from; "bone the turkey before roasting it")

WordNet 2.1 Browser

File History Options Help

Search Word: Redisplay Overview

Searches for bones: Senses:

1 sense of bones

Sense 1

{02841411} <noun.artifact> **bones**#1, castanets#1, clappers#1, finger cymbals#1, maraca#1 -- (a percussion instrument consisting of a pair of hollow pieces of wood or bone (usually held between the thumb and fingers) that are made to click together (as by Spanish dancers) in rhythm with the dance)

- => {03871513} <noun.artifact> percussion instrument#1, percussive instrument#1 -- (a musical instrument in which the sound is produced by one object striking another)
- => {03758578} <noun.artifact> musical instrument#1, instrument1#6 -- (any of various devices or contrivances that can be used to produce musical tones or sounds)
- => {03150574} <noun.artifact> device#1 -- (an instrumentality invented for a particular purpose; "the device is small enough to wear on your wrist"; "a device intended to conserve water")
- => {03536069} <noun.artifact> instrumentality#3, instrumentation#1 -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
- => {00020846} <noun.Tops> artifact#1, artefact#1 -- (a man-made object taken as a whole)
- => {00003991} <noun.Tops> whole#2, unit#6 -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")
- => {00003122} <noun.Tops> object#1, physical object#1 -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
- => {00001930} <noun.Tops> physical entity#1 -- (an entity that has physical existence)
- => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

3 senses of bone

Sense 1

{05206933} <noun.body> **bone**#1, os#5 -- (rigid connective tissue that makes up the skeleton of vertebrates)

- => {05223460} <noun.body> connective tissue#1 -- (tissue of mesodermal origin consisting of e.g. collagen fibroblasts and fatty cells; supports organs and fills spaces between them and forms tendons and ligaments)
- => {05204616} <noun.body> animal tissue#1 -- (the tissue in the bodies of animals)
- => {05204413} <noun.body> tissue#1 -- (part of an organism consisting of an aggregate of cells having a similar structure and function)
- => {05158749} <noun.body> body part#1 -- (any part of an organism such as an organ or extremity)
- => {09251280} <noun.object> part#7, piece#3 -- (a portion of a natural object; "they analyzed the river into three parts"; "he needed a piece of granite")
- => {00002533} <noun.Tops> thing#12 -- (a separate and self-contained entity)
- => {00001930} <noun.Tops> physical entity#1 -- (an entity that has physical existence)
- => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 2

{14563523} <noun.substance> **bone**#2, osseous tissue#1 -- (the porous calcified substance from which bones are made)

- => {14561479} <noun.substance> animal material#1 -- (material derived from animals)
- => {14387488} <noun.substance> material#1, stuff#1 -- (the tangible substance that goes into the makeup of a physical object; "coal is a hard black material"; "wheat is

"Hypernyms (this is a kind of...)" search for noun "bones"

start | Inbox - M... | 2 Windo... | Microsoft ... | Microsoft ... | 2 Micros... | Arthritis -... | WordNet ... | 2:46 PM

Search Word: bones

Redisplay Overview

Searches for bones: Noun Verb

Senses:

1 of 3 senses of bone

Sense 1

- (05206933) <noun.body> **bone**#1, os#5 -- (rigid connective tissue that makes up the skeleton of vertebrates)
 - => (01876081) <noun.animal> furcula#1 -- (a forked bone formed by the fusion of the clavicles of most birds)
 - => (01876237) <noun.animal> wishbone#1, wishing bone#1 -- (the furcula of a domestic fowl)
 - => (02301073) <noun.animal> splint bone#1 -- (a rudimentary metacarpal or metatarsal bone on either side of the cannon bone in the leg of a horse or related animal)
 - => (02439696) <noun.animal> pastern#1, fetter bone#1 -- (the part between the fetlock and the hoof)
 - => (02440654) <noun.animal> cannon bone#1 -- (greatly developed metatarsal or metacarpal bone in the shank or cannon part of the leg in hoofed mammals)
 - => (02445064) <noun.animal> fishbone#1 -- (a bone of a fish)
 - => (05208415) <noun.body> anklebone#1, astragal#1, astragalus#2, talus#2 -- (the bone in the ankle that articulates with the leg bones to form the ankle joint)
 - => (05208639) <noun.body> bare bone#1 -- (bone stripped of flesh)
 - => (05208717) <noun.body> cuboid bone#1 -- (the cube shaped bone on the outer side of the tarsus)
 - => (05208846) <noun.body> carpal bone#1, carpal#1, wrist bone#1 -- (any of the eight small bones of the wrist of primates)
 - => (05209308) <noun.body> scaphoid bone#1, os scaphoideum#1, navicular#1 -- (the largest wrist bone on the thumb side)
 - => (05209437) <noun.body> lunate bone#1, semilunar bone#1, os lunatum#1 -- (one of the eight small wrist bones)
 - => (05209559) <noun.body> triquetral#1, triquetral bone#1, os triquetrum#1, cuneiform bone#1, pyramidal bone#1 -- (a wrist bone that articulates with the pisiform and hamate and lunate bones)
 - => (05209759) <noun.body> pisiform#1, pisiform bone#1, os pisiforme#1 -- (a small wrist bone that articulates only with the triquetral)
 - => (05209905) <noun.body> trapezium#3, trapezium bone#1, os trapezium#1 -- (the wrist bone on the thumb side of the hand that articulates with the 1st and 2nd metacarpals)
 - => (05210087) <noun.body> trapezoid#2, trapezoid bone#1, os trapezoideum#1 -- (the wrist bone between the trapezium and the capitate bones)
 - => (05210237) <noun.body> capitate#1, capitate bone#1, os capitatum#1 -- (the wrist bone with a rounded head shape that articulates with the 3rd metacarpus)
 - => (05210404) <noun.body> hamate#1, hamate bone#1, unciform bone#1, os hamatum#1 -- (the wrist bone in line with the 4th and 5th fingers)
 - => (05210551) <noun.body> cartilage bone#1 -- (any bone that develops within cartilage rather than a fibrous tissue)
 - => (05210680) <noun.body> centrum#1 -- (the main body of a vertebra)
 - => (05210818) <noun.body> cheekbone#1, zygomatic bone#1, zygomatic#1, malar#1, malar bone#1, jugal bone#1, os zygomaticum#1 -- (the arch of bone beneath the eye that forms the prominence of the cheek)
 - => (05211101) <noun.body> clavicle#1, collarbone#1 -- (bone linking the scapula and sternum)
 - => (05211243) <noun.body> coccyx#1, tail bone#1 -- (the end of the vertebral column in humans and tailless apes)
 - => (05211442) <noun.body> dentine#2, dentin#2 -- (bone (calcified tissue) surrounding the pulp cavity of a tooth)
 - => (05211586) <noun.body> ethmoid#1, ethmoid bone#1 -- (one of the eight bones of the cranium; a small bone filled with air spaces that forms part of the eye sockets and the nasal cavity)
 - => (05211804) <noun.body> heelbone#1, calcaneus#1, os tarsi fibulare#1 -- (the largest tarsal bone; forms the human heel)
 - => (05211955) <noun.body> hipbone#1, innominate bone#1 -- (large flaring bone forming one half of the pelvis; made up of the ilium and ischium and pubis)

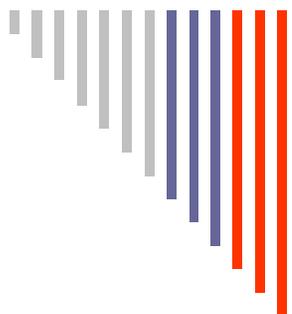
"Hyponyms (... is a kind of this), full" search for noun "bones"



WordNet: manual exploration

Manual extraction – from a single consumer health test document

- 148 medical noun phrases identified
- 52 noun phrases not covered, including
 - 1 drug name - sulfasalazine
 - 1 drug category acronym – DMARD
 - 50 compound noun phrases
- WordNet yielded 96 exact matches for a total noun phrase coverage of ~ 65%

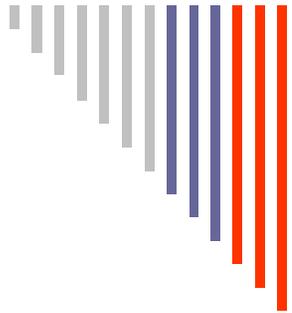


Manual synonym search

For the 50 compound noun phrases with no exact string match, we found:

Exact synonym -	4
Close synonym -	8
Partial synonym -	6
No synonym -	32

Adjusted coverage: ~ 73%



Automated term extraction

PhrasEx – shallow syntactic parsing

Example sentence in document:

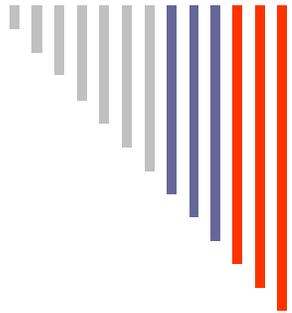
“Your hip is a ball and socket joint, formed by the upper end of the femur, the ball, and a part of the pelvis called the acetabulum, the socket.”

Resulting extracted phrases:

10 simp - “socket”, “acetabulum”, “pelvis”, “part”, “ball”, “femur”, “upper end”, “formed”, “ball and socket joint”, “hip.”

2 macro - “part of pelvis”, “formed by upper end of femur”

3 mega - “hip ball and socket joint formed by upper end of femur ball and part of pelvis”, “acetabulum socket”



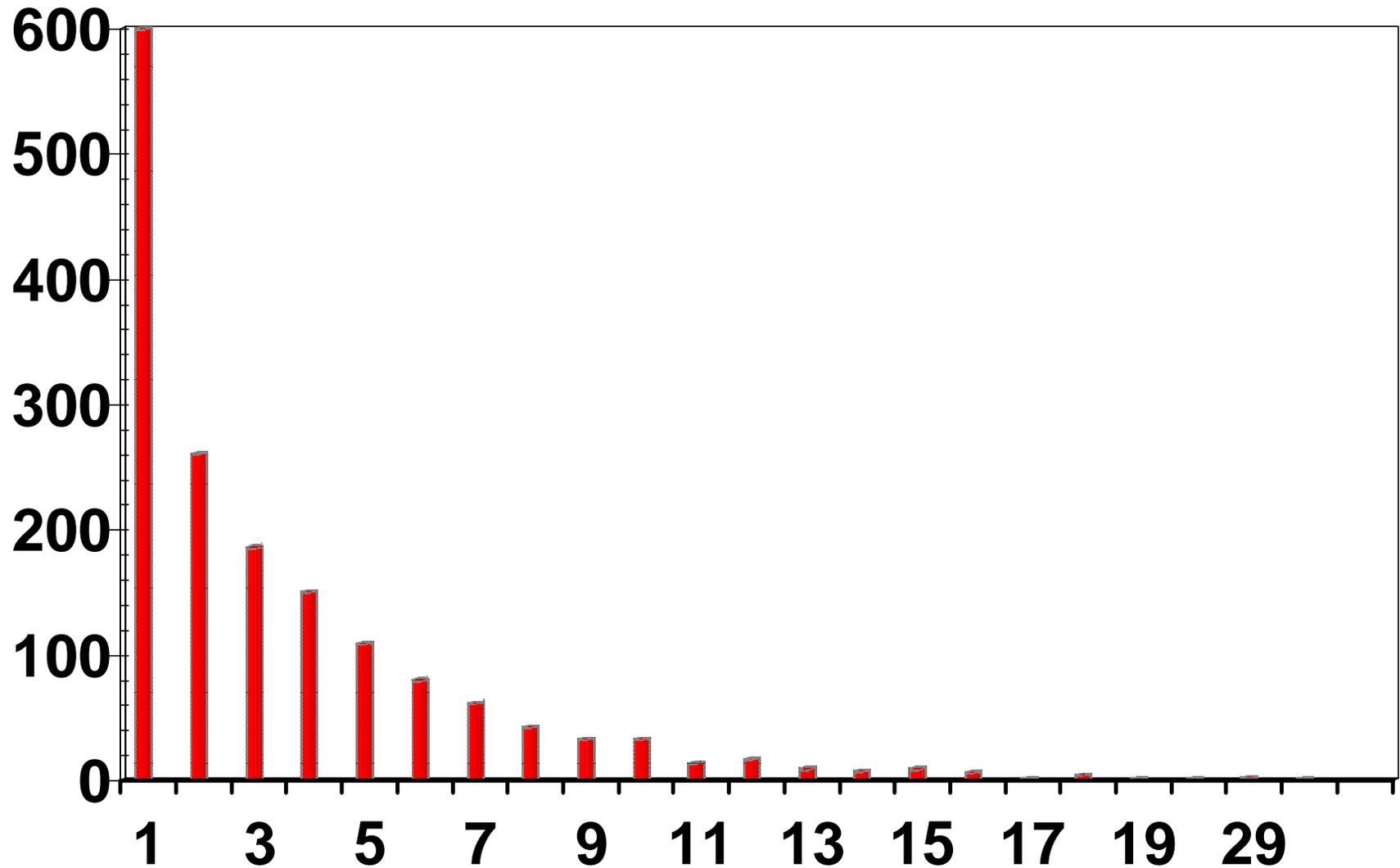
Noun phrase extraction results

- 15,107 phrases extracted (PhrasEx artificially inflates tokens)
- 13,770 actual tokens – occurrences of phrases
- 8,916 types – unique phrases
 - **Simp** types – 4,678
 - **Macro** types – 1,621
 - **Mega** types – 2,617

Initial mapping of extracted phrases to WordNet*

(*WN 2.0)	No WN mapping	Some WN mapping
Total tokens	8,581 (62%)	5,189 (38%)
simp	4,003	5,140
macro	1,708	4
mega	2,870	45
Total types	7,282 (82%)	1,634 (18%)
simp	3,074	1,604
macro	1,617	4
mega	2,591	26

of synsets mapped to per type



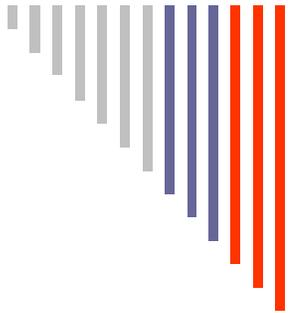


Top 15 most frequent phrases mapped to something in WN

- 169|simp|arthritis|n|13380097%n|arthritis|1|0
- 131|simp|pain|n|13510751%n|pain//hurting|5|0
- 75|simp|joints|n|05275200%n|joint//articulation//articulatio|6|0
- 73|simp|people|n|07463651%n|people|4|0
- 66|simp|i|n|13803148%n|iodine//iodin//I//atomic number 53|3|0
- 64|simp|rheumatoid arthritis|n|13380258%n|rheumatoid arthritis//atrophic arthritis//rheumatism|1|0
- 59|simp|joint|n|05275200%n|joint//articulation//articulatio|6|0
- 54|simp|patients|n|09730887%n|patient|2|0
- 49|simp|body|n|04916110%n|body//organic structure//physical structure|9|0
- 48|simp|ra|n|13813213%n|radium//Ra//atomic number 88|3|0
- 48|simp|osteoarthritis|n|13381668%n|osteoarthritis//degenerative arthritis//degenerative joint disease|1|0
- 44|simp|doctor|n|09380179%n|doctor//doc//physician//MD//Dr//medico|4|0
- 42|simp|inflammation|n|13524046%n|inflammation//redness//rubor|4|0
- 38|simp|disease|n|13268088%n|disease|1|0
- 34|simp|cartilage|n|04983662%n|cartilage//gristle|1|0

Initial mapping of extracted phrases to the UMLS*

(*UMLS2005AA)	No Metathesaurus mapping	Some Metathesaurus mapping
Total tokens	7786 (57%)	5984 (43%)
simp	3,501	5,642
macro	1,571	141
mega	2,714	201
Total types	6796 (76%)	2120 (24%)
simp	2,826	1,852
macro	1,503	118
mega	2,467	150



Categorization/Disambiguation

Our semantic categories for disambiguation fit 292 out of 1,634 types mapped to at least one WordNet synset.

- 291 types – relevant to 1 category
One exception - “joint” is relevant to 2 categories

192 of these cases were mapped to 1 synset only identified as being of biomedical interest

In the other 99 cases with multiple mappings, the semantic categories identify which of the mapped-to synsets are biomedical.



File History Options Help

Search Word:

Searches for tendon: Senses:

1 sense of tendon

Sense 1

tendon, sinew -- (a cord or band of inelastic tissue connecting a muscle with its bony attachment)

- => connective tissue -- (tissue of mesodermal origin consisting of e.g. collagen fibroblasts and fatty cells; supports organs and fills spaces between them and forms tendons and ligaments)
- => animal tissue -- (the tissue in the bodies of animals)
 - => tissue -- (part of an organism consisting of an aggregate of cells having a similar structure and function)
 - => body part -- (any part of an organism such as an organ or extremity)
 - => part, piece -- (a portion of a natural object; "they analyzed the river into three parts"; "he needed a piece of granite")
 - => thing -- (a separate and self-contained entity)
 - => physical entity -- (an entity that has physical existence)
 - => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"Hypernyms (this is a kind of...)" search for noun "tendon"



File History Options Help

Search Word:

Searches for cyst: Senses:

2 senses of cyst

Sense 1

cyst -- (a closed sac that develops abnormally in some body structure)

- => pathology -- (any deviation from a healthy or normal condition)
- => ill health, unhealthiness, health problem -- (a state in which you are unable to function normally and without pain)
- => pathological state -- (a condition that is not normal)
- => condition, status -- (a state at a particular time; "a condition (or state) of disrepair"; "the current status of the arms negotiations")
- => state -- (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state")
- => attribute -- (an abstraction belonging to or characteristic of an entity)
- => abstraction -- (a general concept formed by extracting common features from specific examples)
- => abstract entity -- (an entity that exists only abstractly)
- => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 2

vesicle, **cyst** -- (a small anatomically normal sac or bladderlike structure (especially one containing fluid))

- => sac -- (a structure resembling a bag in an animal)
- => cavity, bodily cavity, cavum -- ((anatomy) a natural hollow or sinus within the body)
- => structure, anatomical structure, complex body part, bodily structure, body structure -- (a particular complex anatomical part; "he has good bone structure")
- => body part -- (any part of an organism such as an organ or extremity)
- => part, piece -- (a portion of a natural object; "they analyzed the river into three parts"; "he needed a piece of granite")
- => thing -- (a separate and self-contained entity)
- => physical entity -- (an entity that has physical existence)
- => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"Hyponyms (this is a kind of...)" search for noun "cyst"



Discussion

WordNet v. Metathesaurus

1,634 types mapped

(18%)

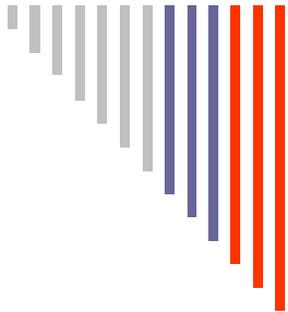
2,120 types mapped

(24%)

Phrases mapped in UMLS but not WN:

9|simp| [fibromyalgia](#) |EM|1|C0016053|Fibromyalgia

35|simp| [rheumatic diseases](#) |EM|1|C0035435|Rheumatism



Future work...

Expand the document corpus

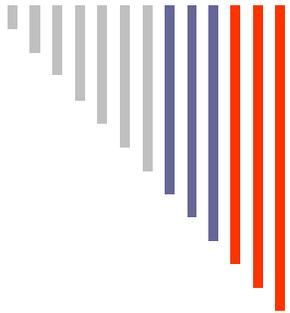
- Increase the number of articles included
- Increase variety of MMV sources, MMV genres, and targeted audiences

Normalize for plurals up front

Expand the disambiguation effort

- Verify and complete our current semantic categories
- Identify more semantic categories

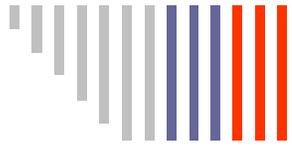
and...



...More future work

Deal with “silence” – automate WordNet alternate-string synonym search

- Join the WordNet and UMLS mappings
- For phrases that do not map to WordNet but map to UMLS, identify the CUI
- Remap all of alternative-string synonyms of those CUIs to WordNet



Learning objectives successfully met:

- To explore and understand the terms, concepts and structuring of the electronic lexical database WordNet.
- To explore and understand Metathesaurus terms and concepts, the semantic types, and the structuring of UMLS.
- To learn how to use UMLS tools to extract terms from documents and map them to the Metathesaurus and to WordNet.
- To learn more about the current state of CHI controlled vocabularies.
- To acquire skills in vocabulary analysis and CHI vocabulary development.
- To explore WordNet as a source for CHI coverage of the domain of arthritis and of medicine/diseases in general.



Additional thank yous...

May Cheh

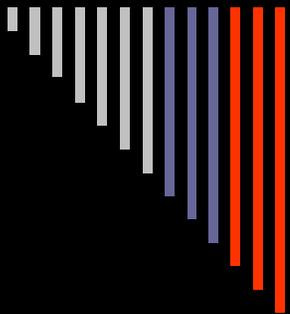
Tom Rindflesch

Allen Browne

Lan Aronson

Rob Logan

**And all of the LHCBC staff for their assistance,
expertise, kindness and time.**



Comments & Questions?

