ihtsdo Leading healthcare terminology, worldwide

SNOMED CT search techniques

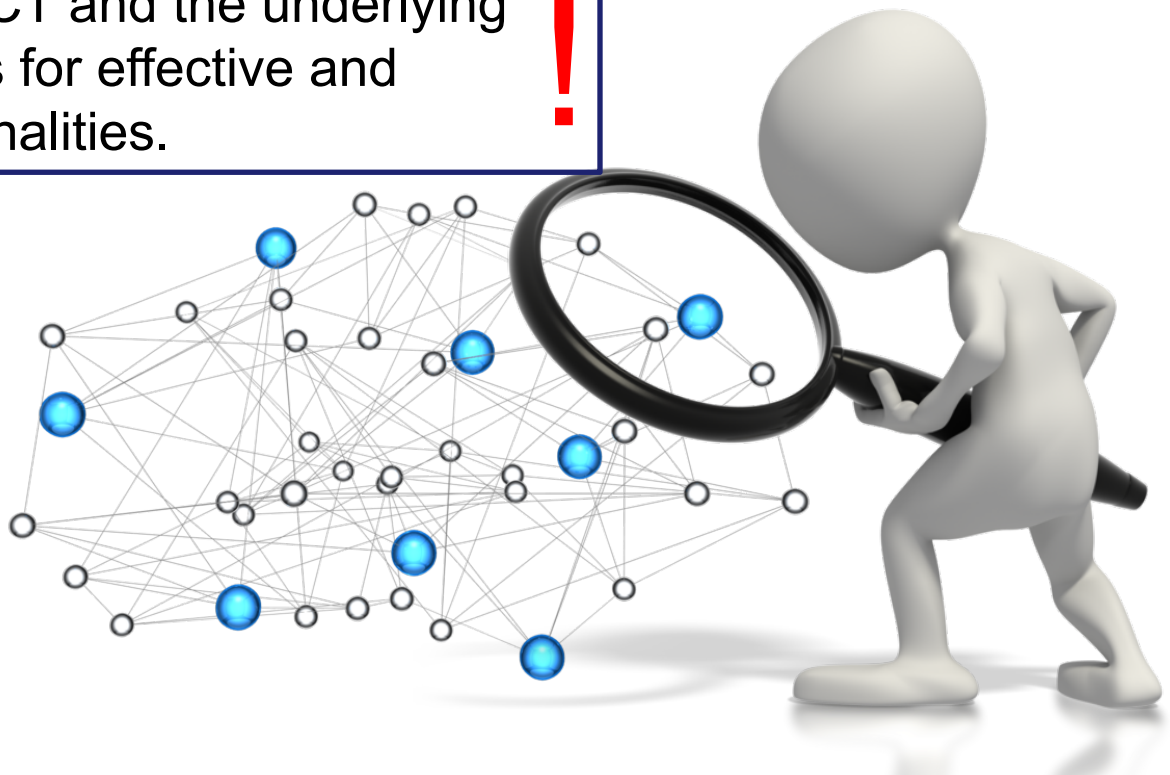
Anne Randorff Højen
M.Sc. Phd.
SNOMED CT Implementation Advisor

Delivering SNOMED CT
The global language of healthcare

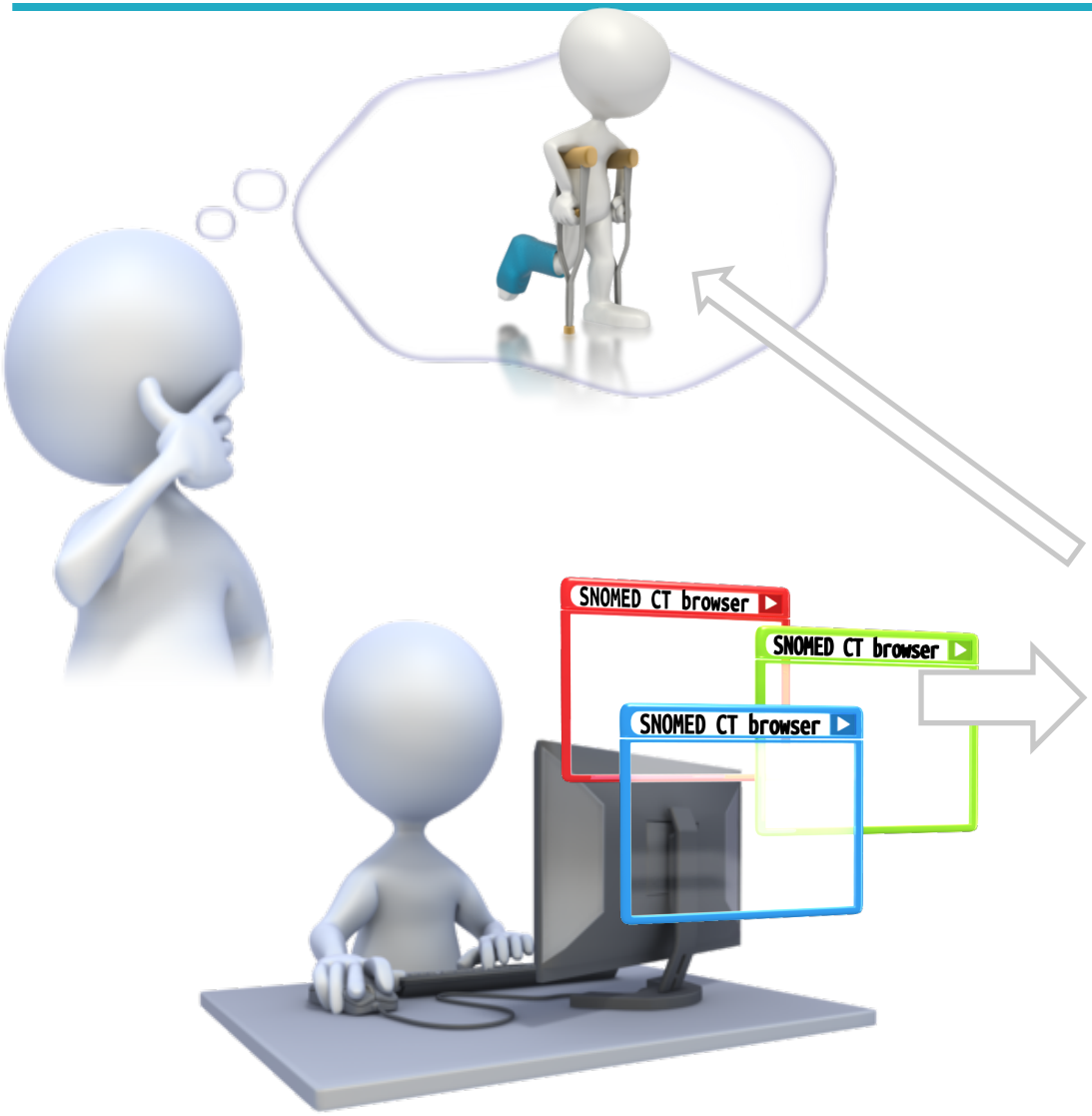
Why do we need search capabilities?

SNOMED CT is the most comprehensive, clinical healthcare terminology in the world.
It includes concepts representing the wide range of types of information that need to be recorded in clinical records.

The size of SNOMED CT and the underlying logical model demands for effective and efficient search functionalities. !

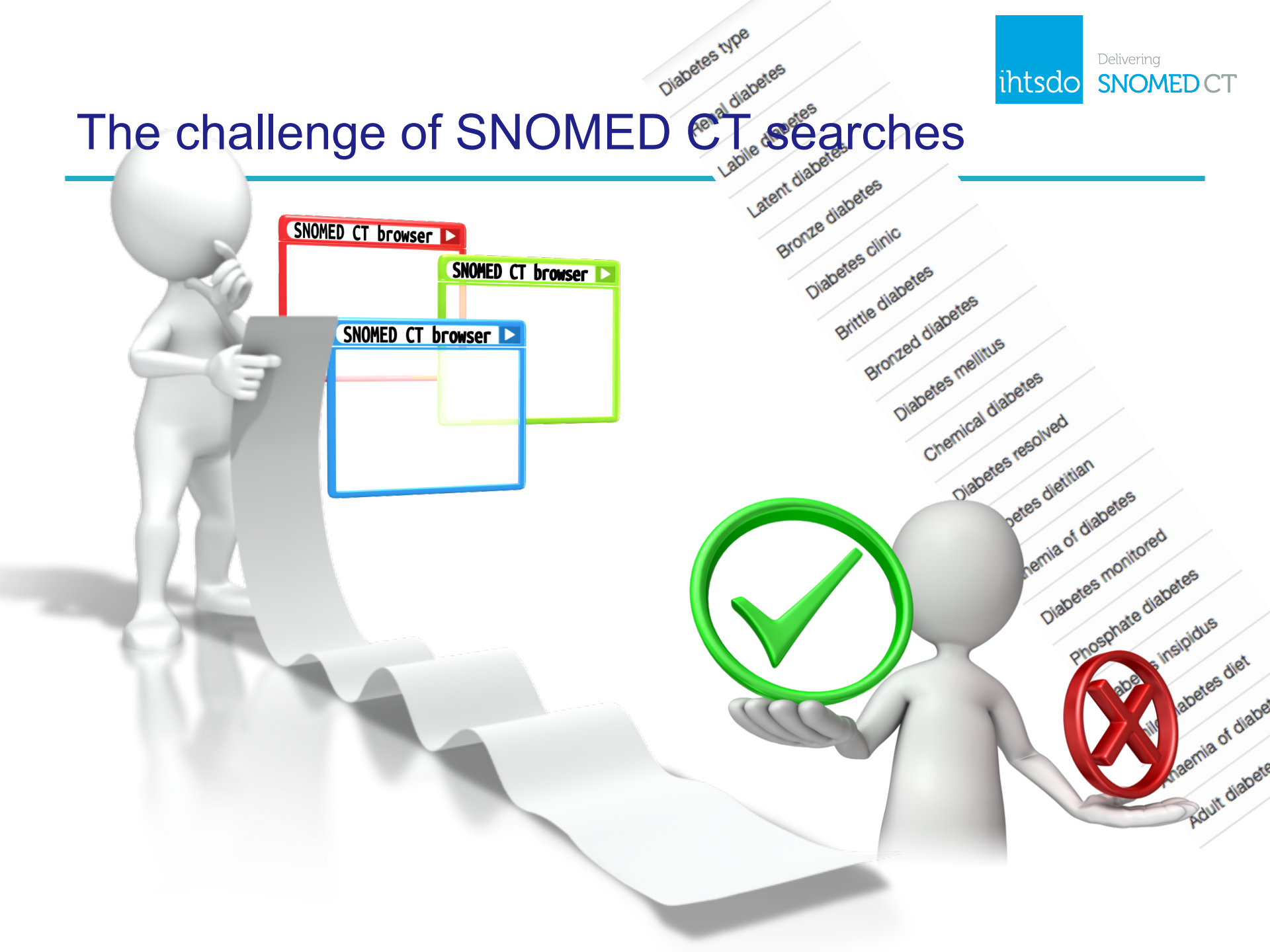


What is SNOMED CT search?



- Fracture of lower limb
- Injury of ankle
- Fracture of ankle
- Closed fracture of ankle
- Fracture dislocation of ankle joint
- Fracture of distal end of fibula
- Fracture of distal end of tibia
- Fracture of talus
- Open fracture of ankle

The challenge of SNOMED CT searches



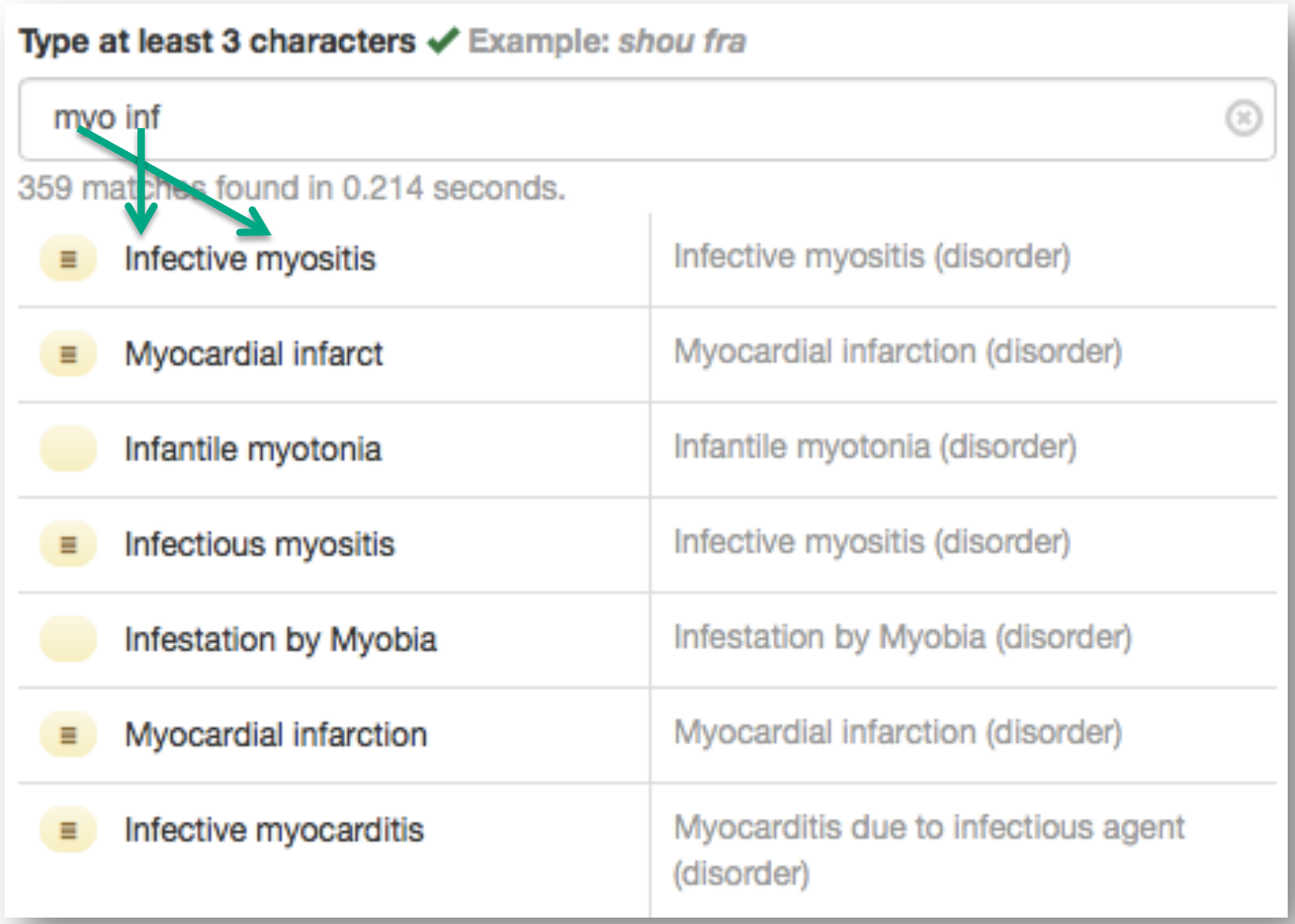
ihtsdo Leading healthcare terminology, worldwide

EXAMPLES OF SEARCH TECHNIQUES

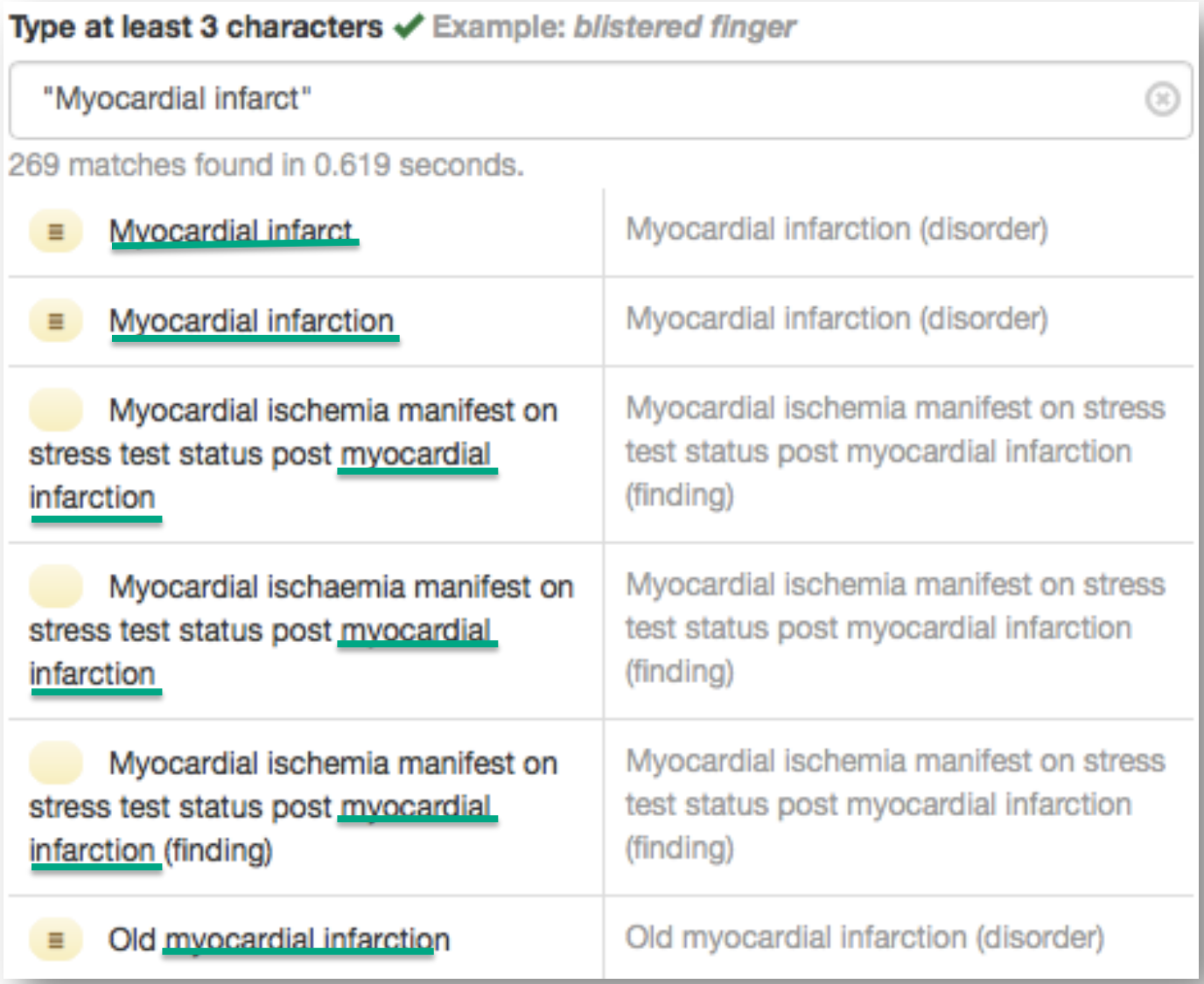
Search by terms

- String searches must be user configurable to support searches, such as searching for:
  - words any order
  - phrase match
  - identical term

Search by terms (Words any order)



Search by terms (Phrase match)



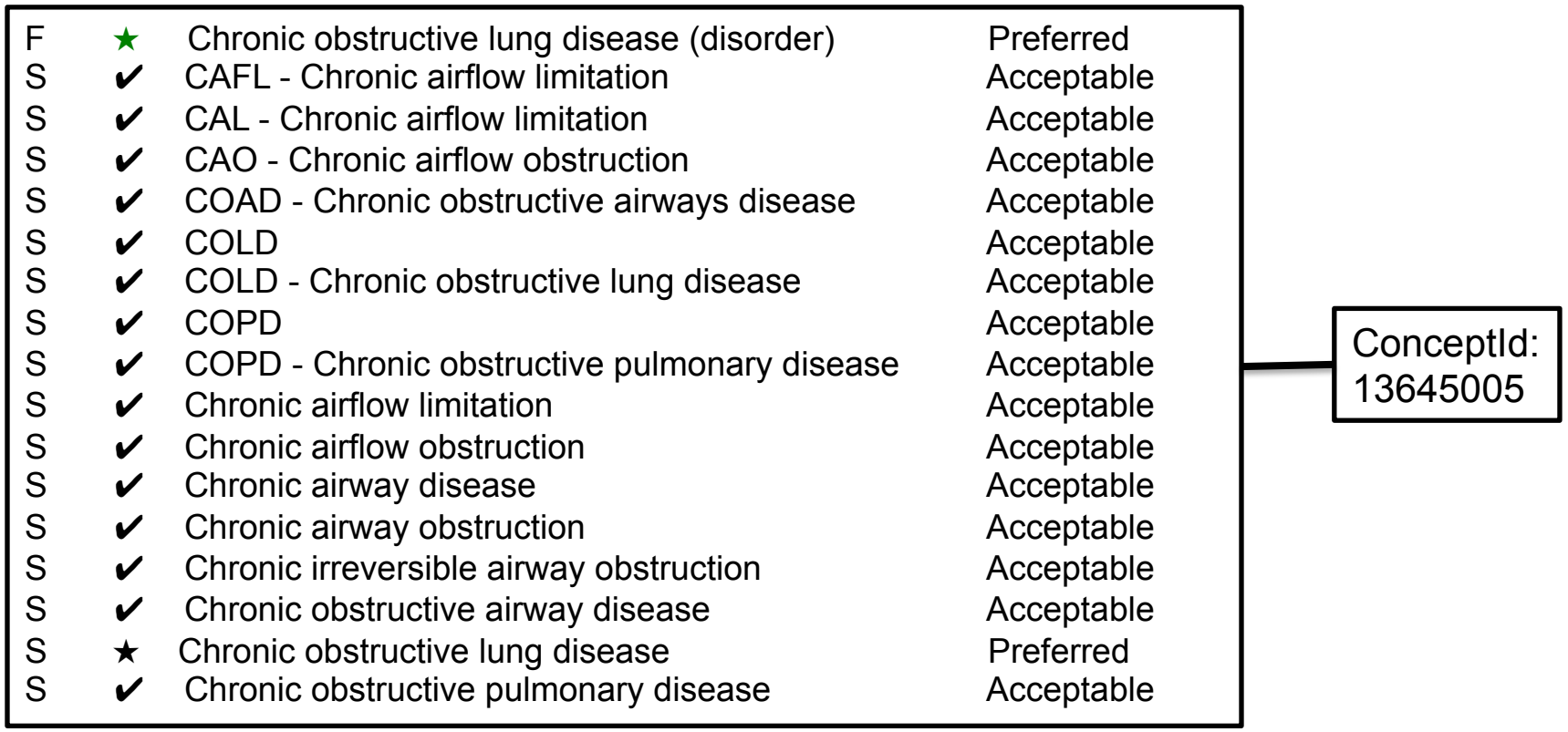
Search by identifiers – Searching for cold (Concept Id:13645005)

ihtsdo Delivering SNOMED CT

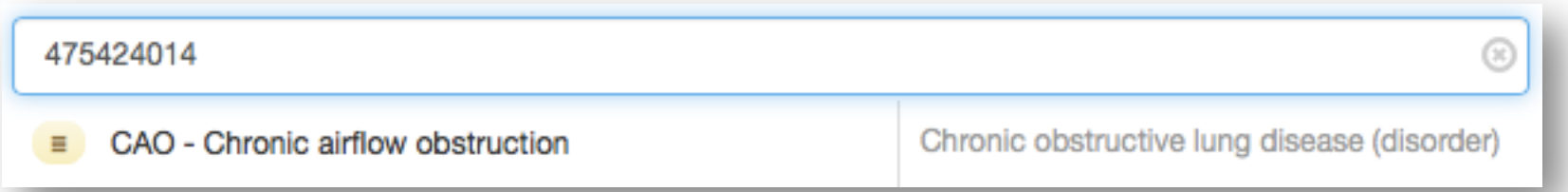
13645005

Chronic obstructive lung disease	Chronic obstructive lung disease (disorder)
COLD	Chronic obstructive lung disease (disorder)
COPD	Chronic obstructive lung disease (disorder)
Chronic airway obstruction	Chronic obstructive lung disease (disorder)
Chronic irreversible airway obstruction	Chronic obstructive lung disease (disorder)
Chronic airflow obstruction	Chronic obstructive lung disease (disorder)
CAFL - Chronic airflow limitation	Chronic obstructive lung disease (disorder)
COAD - Chronic obstructive airways disease	Chronic obstructive lung disease (disorder)
CAO - Chronic airflow obstruction	Chronic obstructive lung disease (disorder)
Chronic airflow limitation	Chronic obstructive lung disease (disorder)
CAL - Chronic airflow limitation	Chronic obstructive lung disease (disorder)
COPD - Chronic obstructive pulmonary disease	Chronic obstructive lung disease (disorder)
Chronic obstructive airway disease	Chronic obstructive lung disease (disorder)
Chronic airway disease	Chronic obstructive lung disease (disorder)
COLD - Chronic obstructive lung disease	Chronic obstructive lung disease (disorder)
Chronic obstructive pulmonary disease	Chronic obstructive lung disease (disorder)
Chronic obstructive lung disease (disorder)	Chronic obstructive lung disease (disorder)

Search by identifiers – Searching for cold
(Concept ID:13645005)



Search by identifiers – Searching for cold (DescriptionId: 475424014)

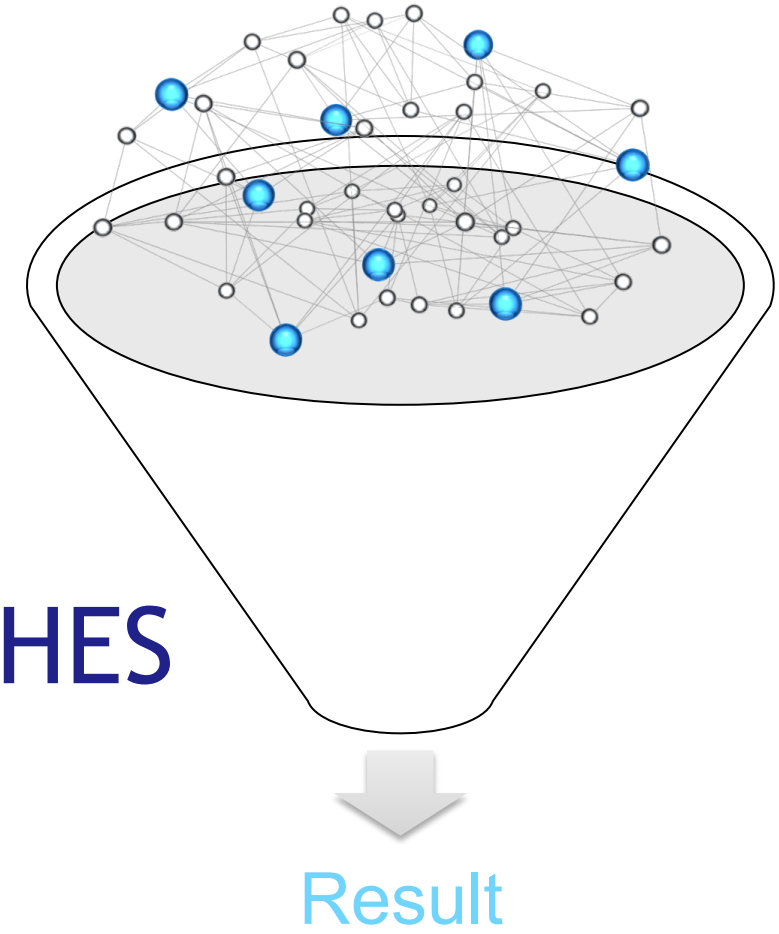


Search by identifiers – Searching for cold
(Concept ID:13645005)

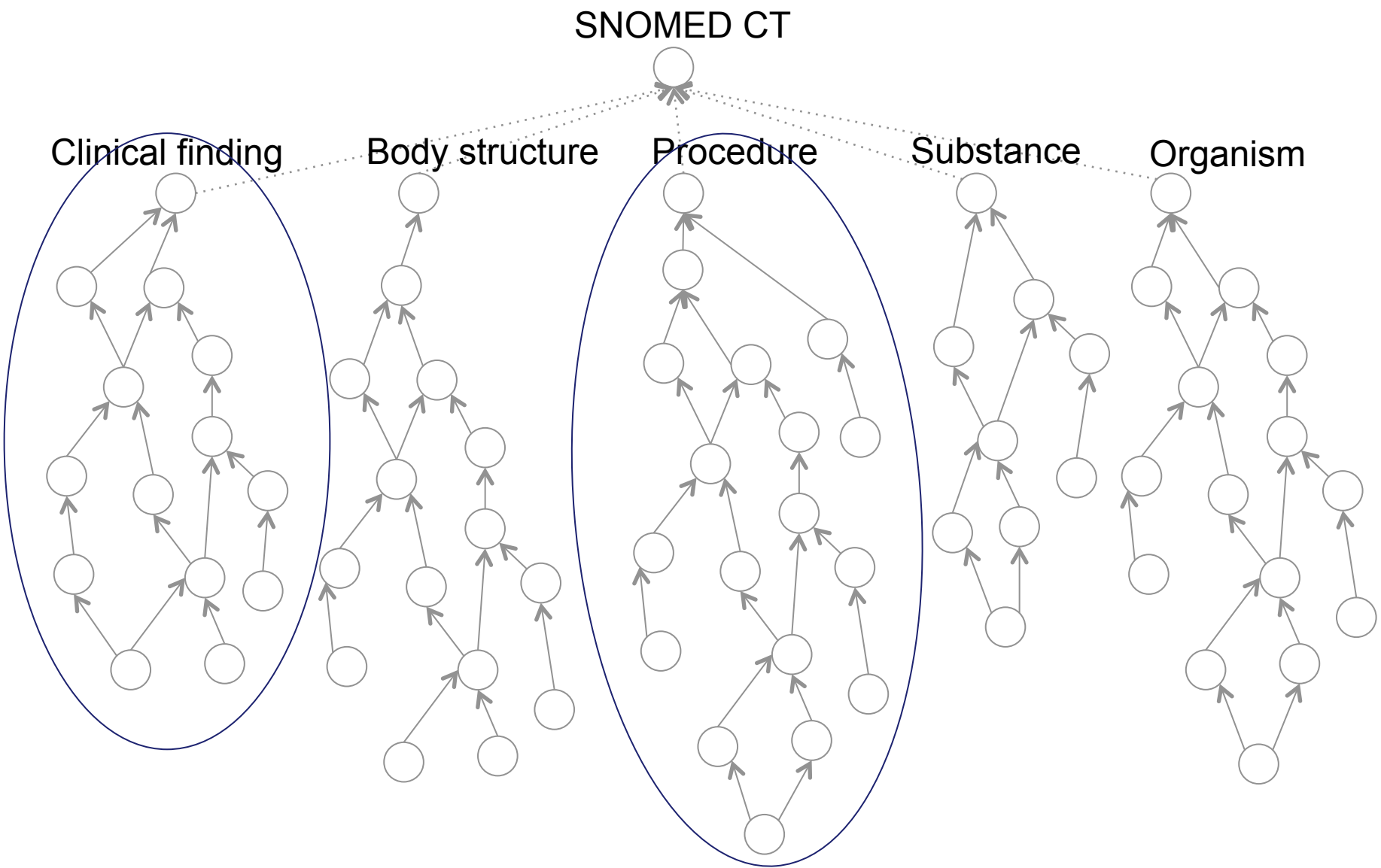
F	★	Chronic obstructive lung disease (disorder)	Preferred	739041012
S	✔	CAFL - Chronic airflow limitation	Acceptable	475422013
S	✔	CAL - Chronic airflow limitation	Acceptable	475426011
S	✔	CAO - Chronic airflow obstruction	Acceptable	475424014
S	✔	COAD - Chronic obstructive airways disease	Acceptable	475423015
S	✔	COLD	Acceptable	23289016
S	✔	COLD - Chronic obstructive lung disease	Acceptable	475430014
S	✔	COPD	Acceptable	23290013
S	✔	COPD - Chronic obstructive pulmonary disease	Acceptable	475427019
S	✔	Chronic airflow limitation	Acceptable	475425010
S	✔	Chronic airflow obstruction	Acceptable	475421018
S	✔	Chronic airway disease	Acceptable	475429016
S	✔	Chronic airway obstruction	Acceptable	475419011
S	✔	Chronic irreversible airway obstruction	Acceptable	475420017
S	✔	Chronic obstructive airway disease	Acceptable	475428012
S	★	Chronic obstructive lung disease	Preferred	23287019
S	✔	Chronic obstructive pulmonary disease	Acceptable	475431013

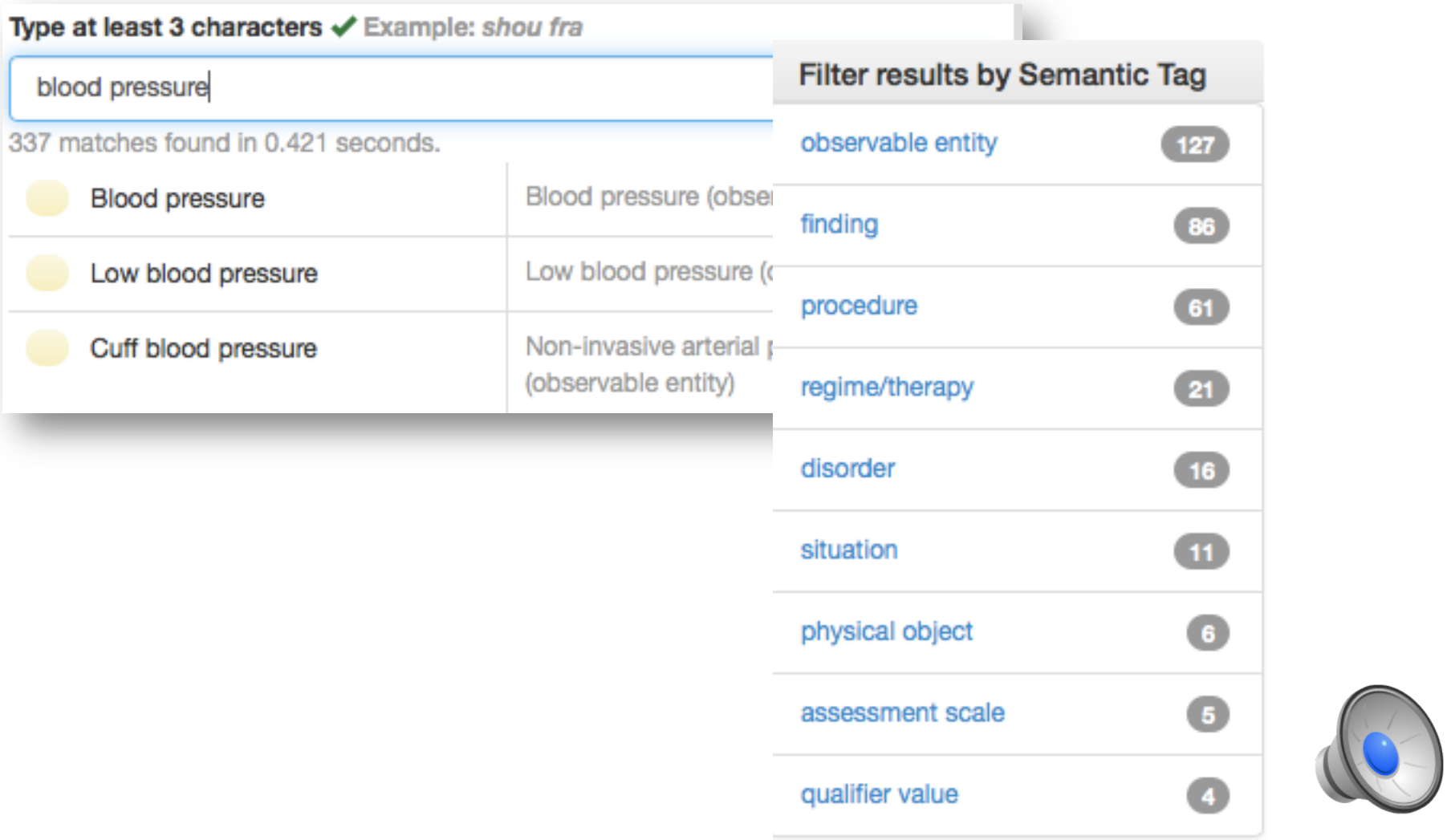
ihtsdo Leading healthcare terminology, worldwide

CONSTRAIN SEARCHES

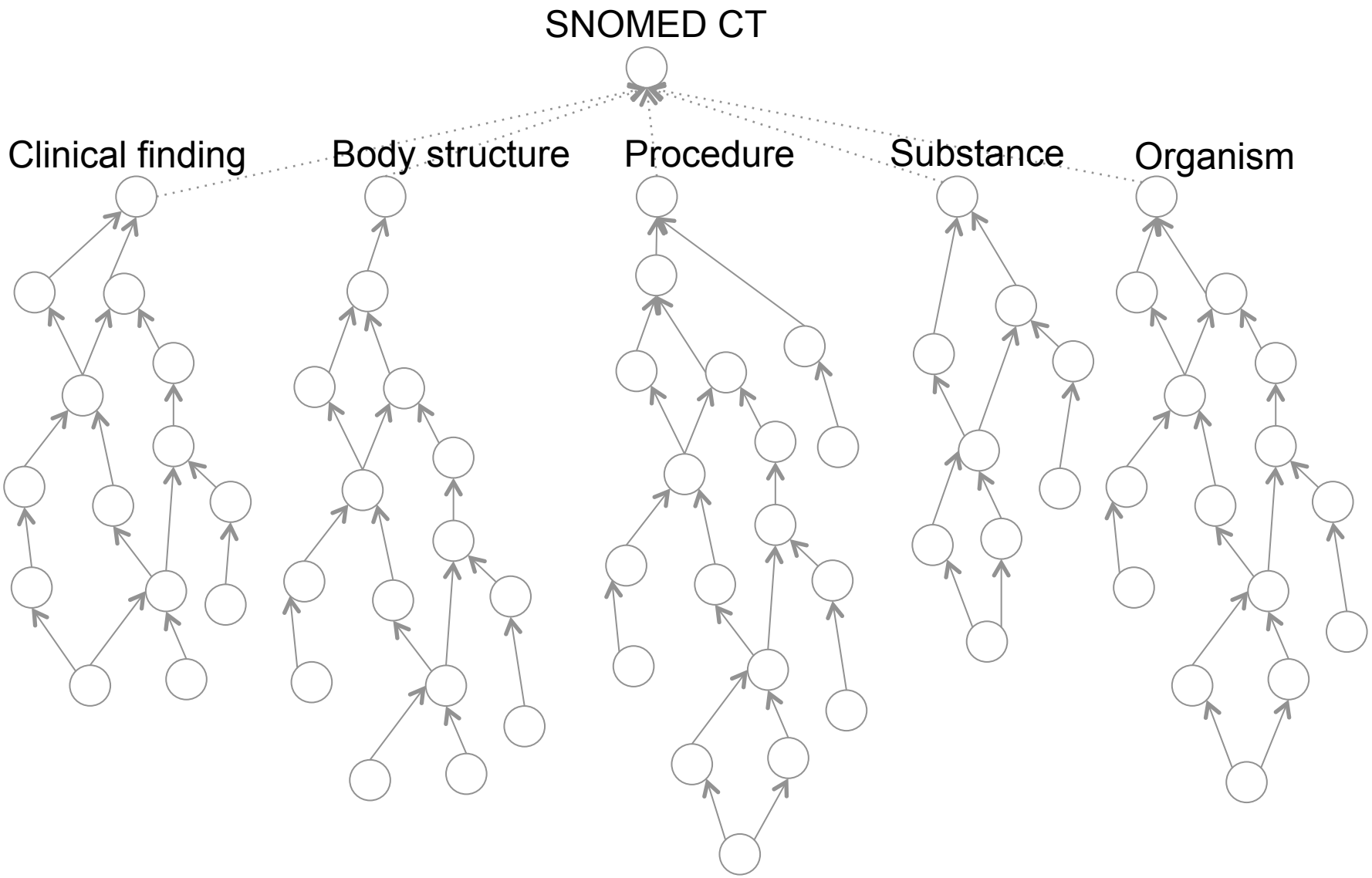


Constrain search by hierarchy

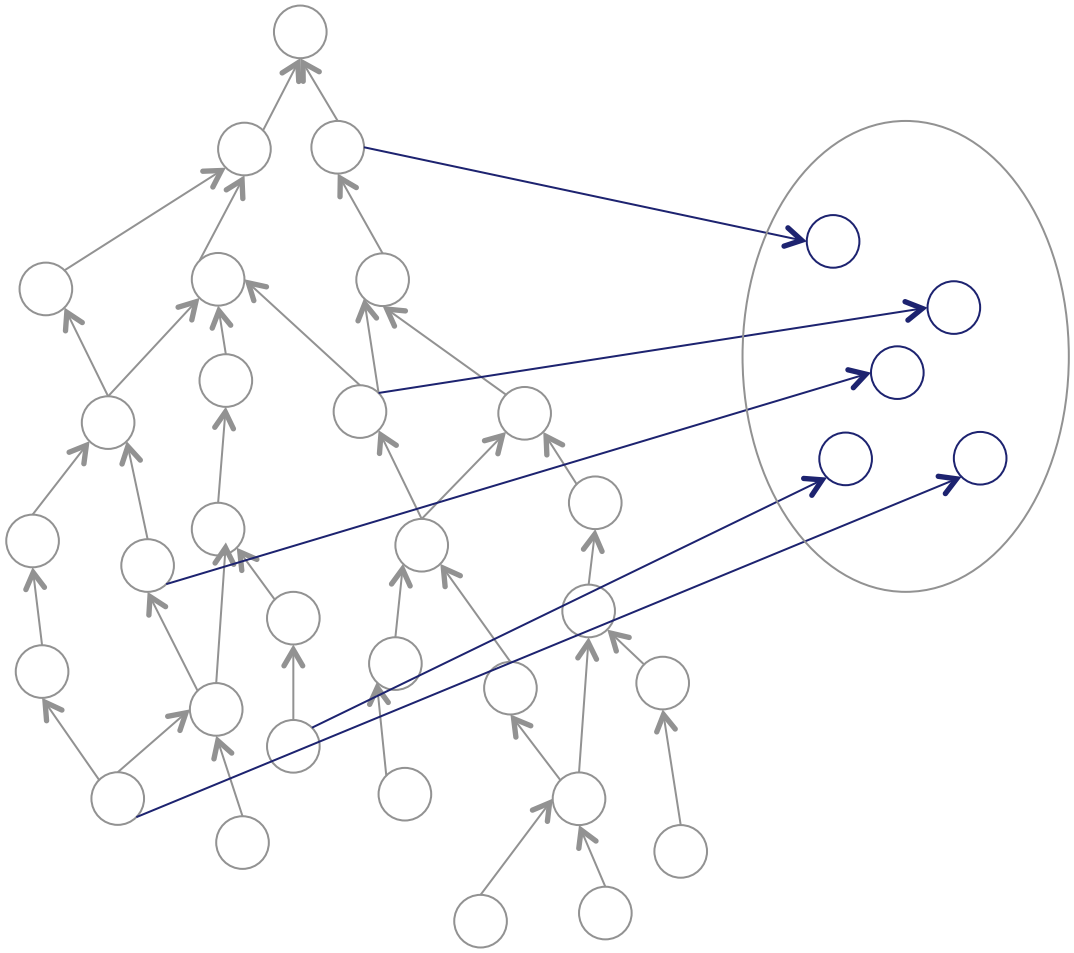



Type at least 3 characters ✔ Example: shou fra
blood pressure
337 matches found in 0.421 seconds.
Blood pressure
Blood pressure (obse
Low blood pressure
Low blood pressure (
Cuff blood pressure
Non-invasive arterial
(observable entity)
Filter results by Semantic Tag
observable entity 127
finding 86
procedure 61
regime/therapy 21
disorder 16
situation 11
physical object 6
assessment scale 5
qualifier value 4

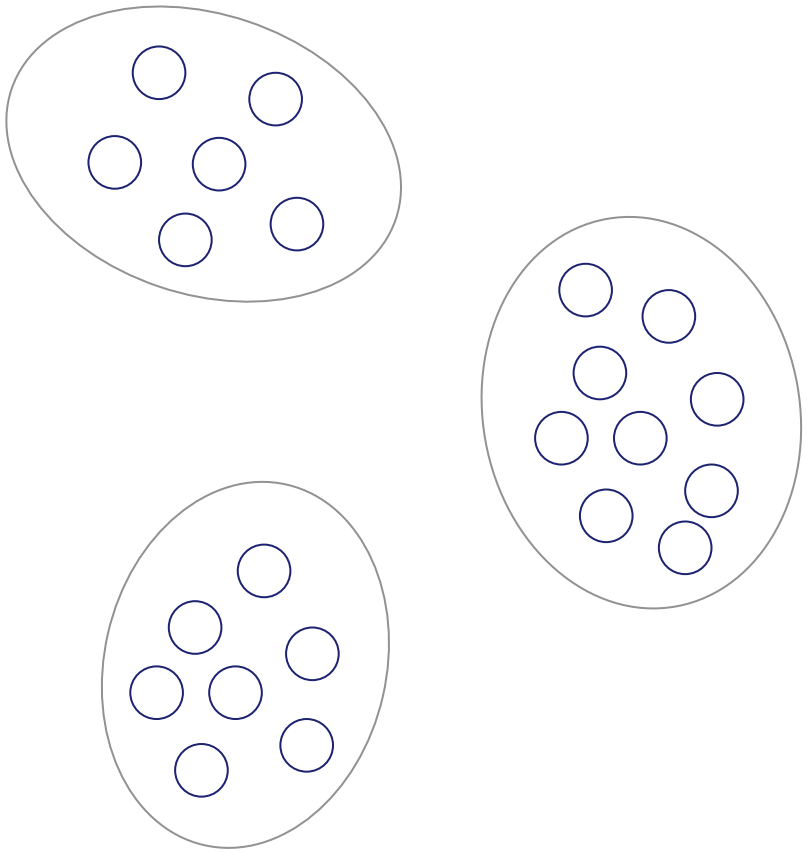
Constrain search by hierarchy



Constrain search and data entry by reference sets



Constrain search by reference sets

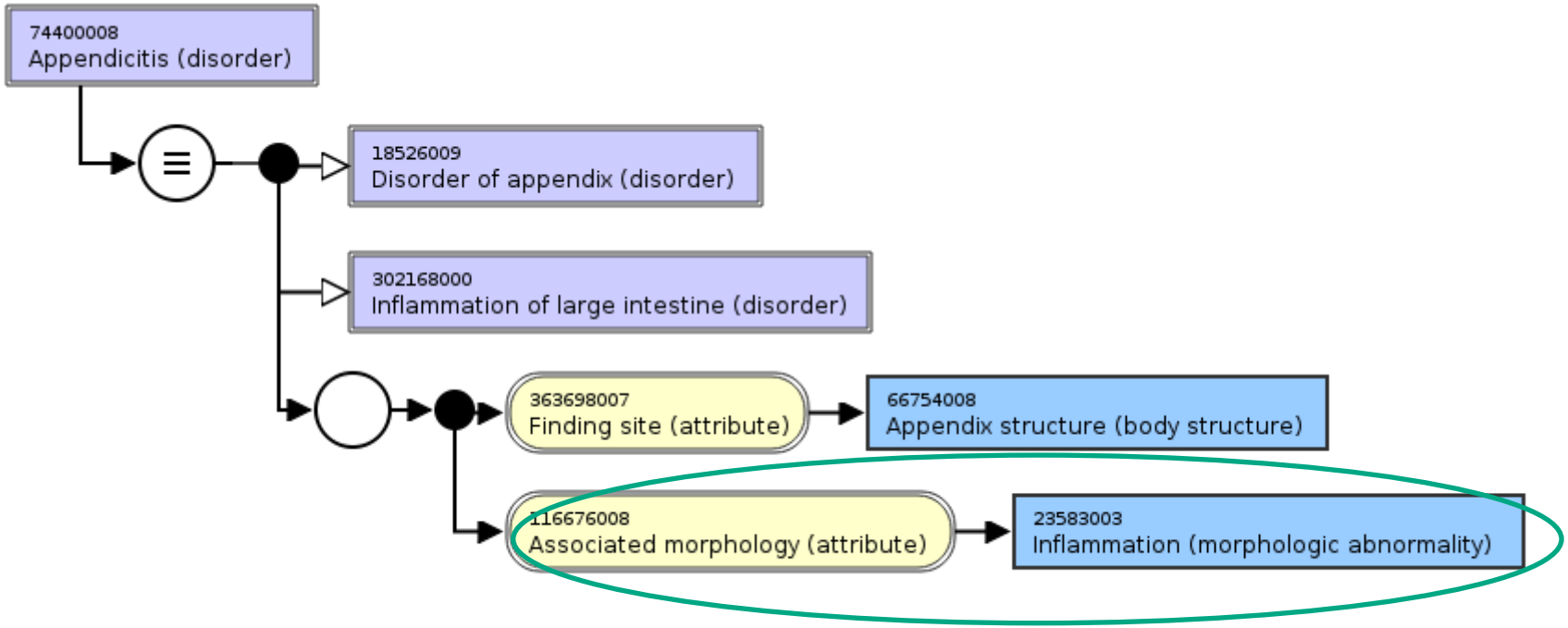


Examples:

- Language ref sets to avoid uncommon or foreign terms
- Simple ref sets to simplify or encourage selection of concepts or used in a particular country, organization, or specialty
- Context ref sets to specify or order the valid Concepts for entry in a particular field

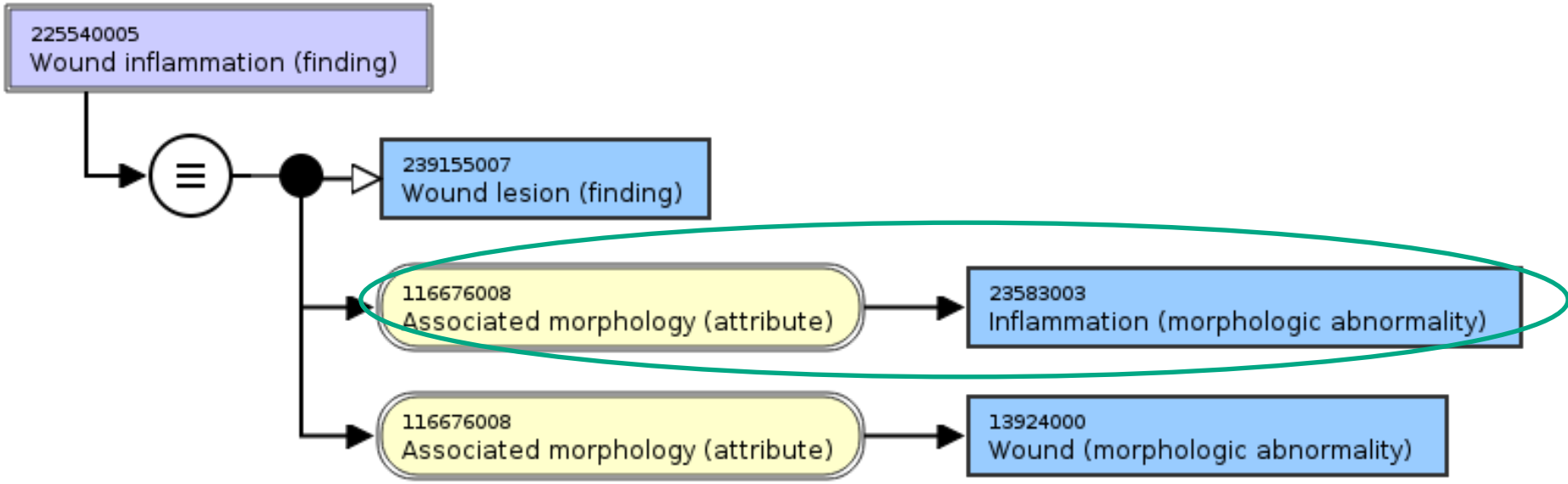
Constrain by attribute

- Constraints based on particular attributes of the concept
  - For example, only concepts with the attribute |associated morphology| having the value |inflammation|.



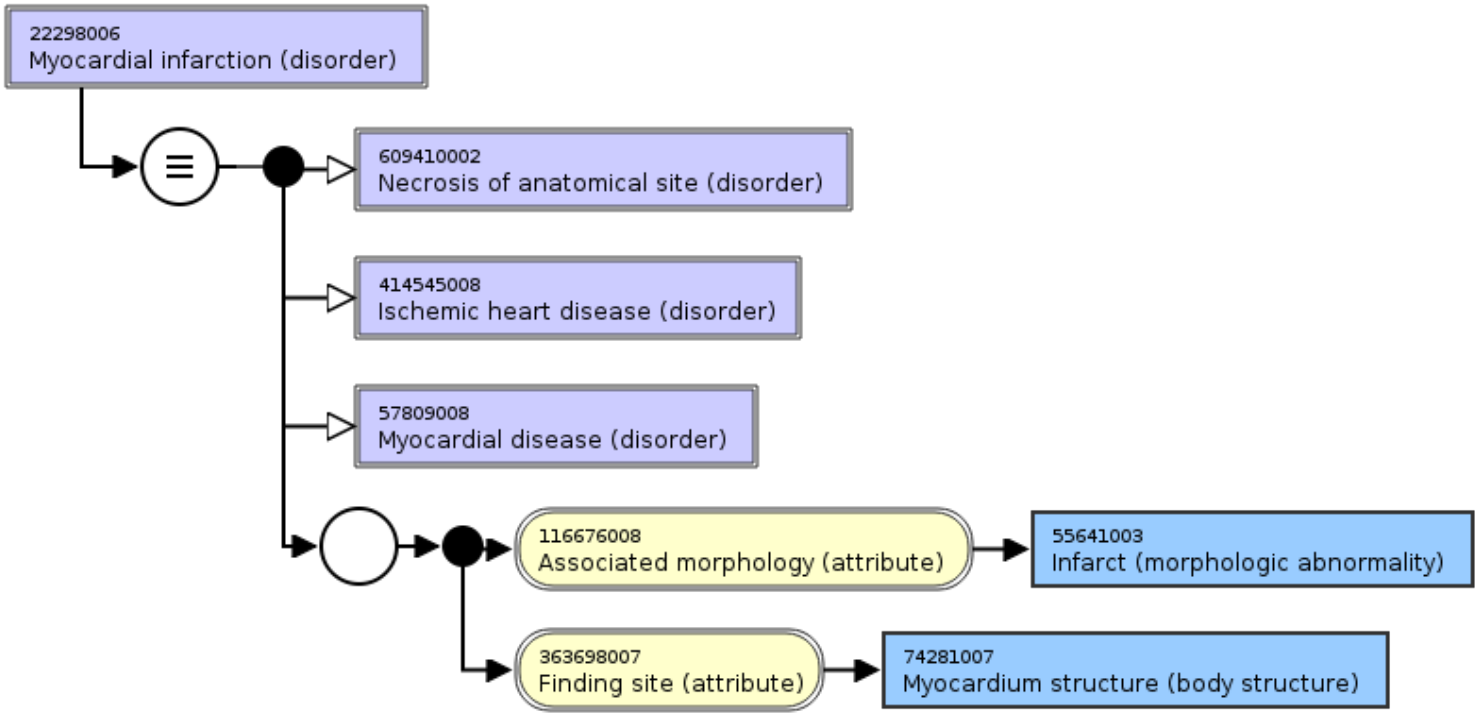
Constrain by attribute

- Constraints based on particular attributes of the concept
  - For example, only concepts with the attribute |associated morphology| having the value |inflammation|.



Constrain by attribute

- Constraints based on particular attributes of the concept
  - For example, only concepts with the attribute |associated morphology| having the value |inflammation|.



ihtsdo Leading healthcare terminology, worldwide

EXTEND SEARCHES

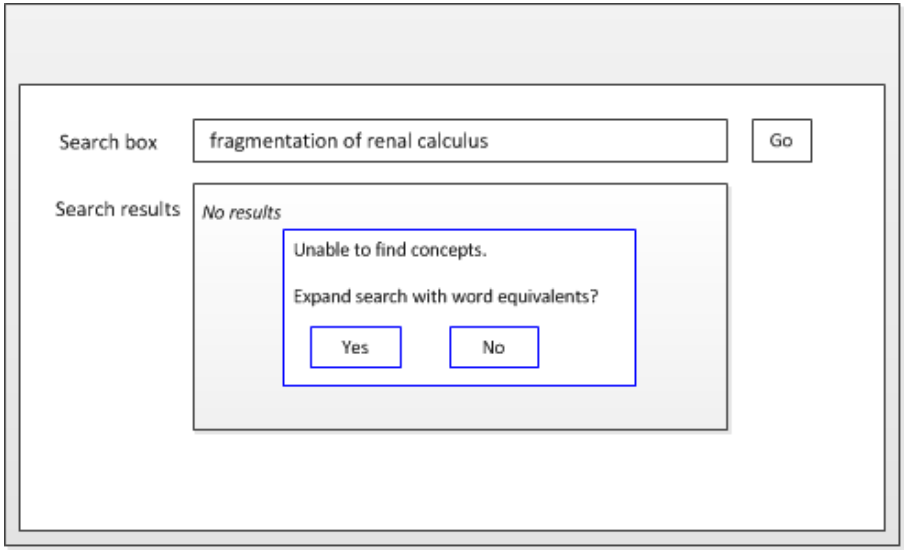
Extend searches – Use Word Equivalents (1)

- In healthcare, there are many words with equivalent meanings
- *Synonyms* provide alternative phrases referring to the *concept*
- *Synonyms* are not created automatically for every possible combination of words with an equivalent meaning

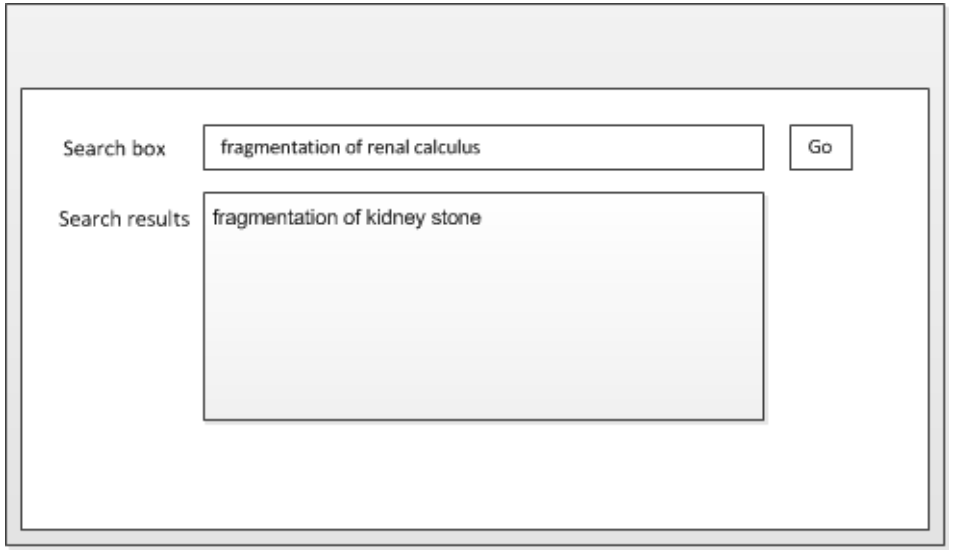
- One way of addressing this problem is to maintain a table of *Word Equivalents*

Extend searches – Use Word Equivalents (2)

1



2



Extend searches: by post-coordinated searching (1)

- When typing text for a search, the user is unlikely to know if their intended entry can be represented by a single Concept or requires a post-coordinated expression involving additional Concepts or qualifiers.

- Where searches fail to find a pre-coordinated match, expansion of the search to support appropriate or commonly used qualifiers is likely to enhance usability.

Extend searches: by post-coordinated searching (2)

ihtsdo Leading healthcare terminology, worldwide

IMPROVING SEARCH SPEEDS

Improve search speeds

- Enable real time searching

- Allow slow searches to be paused or cancelled

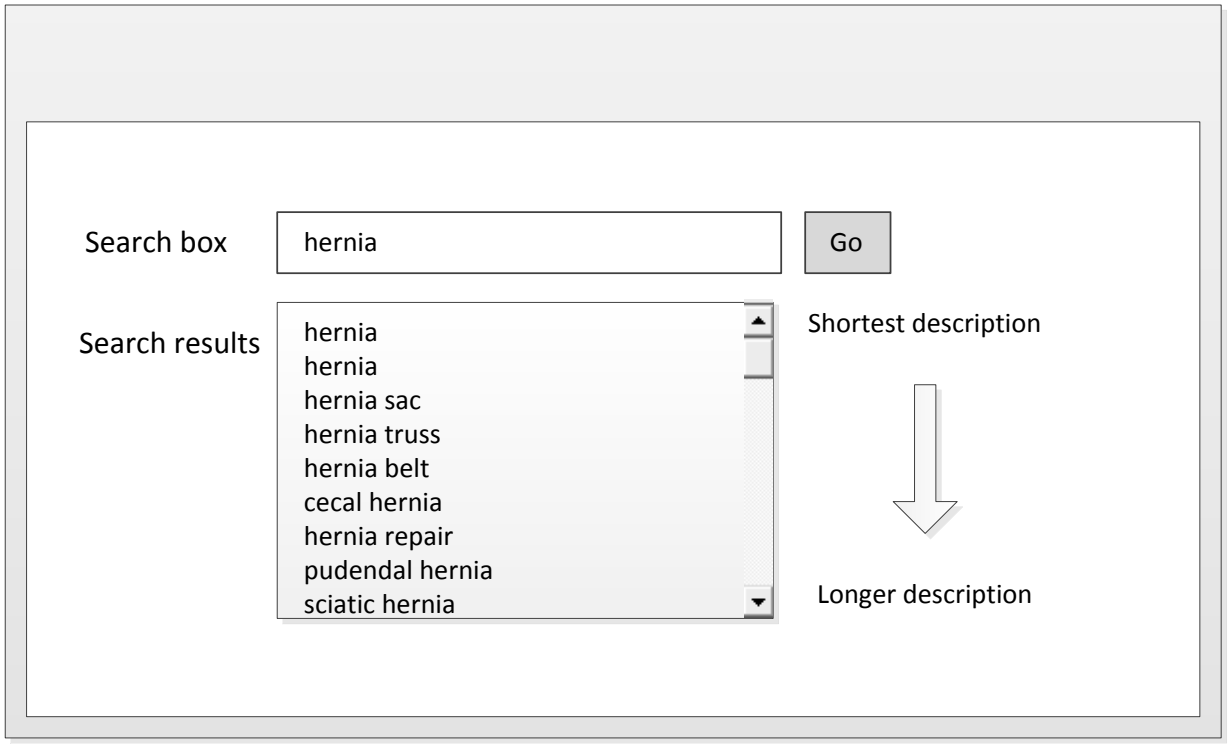
- Optimize indexing

ihtsdo Leading healthcare terminology, worldwide

DISPLAY OF SEARCH RESULTS

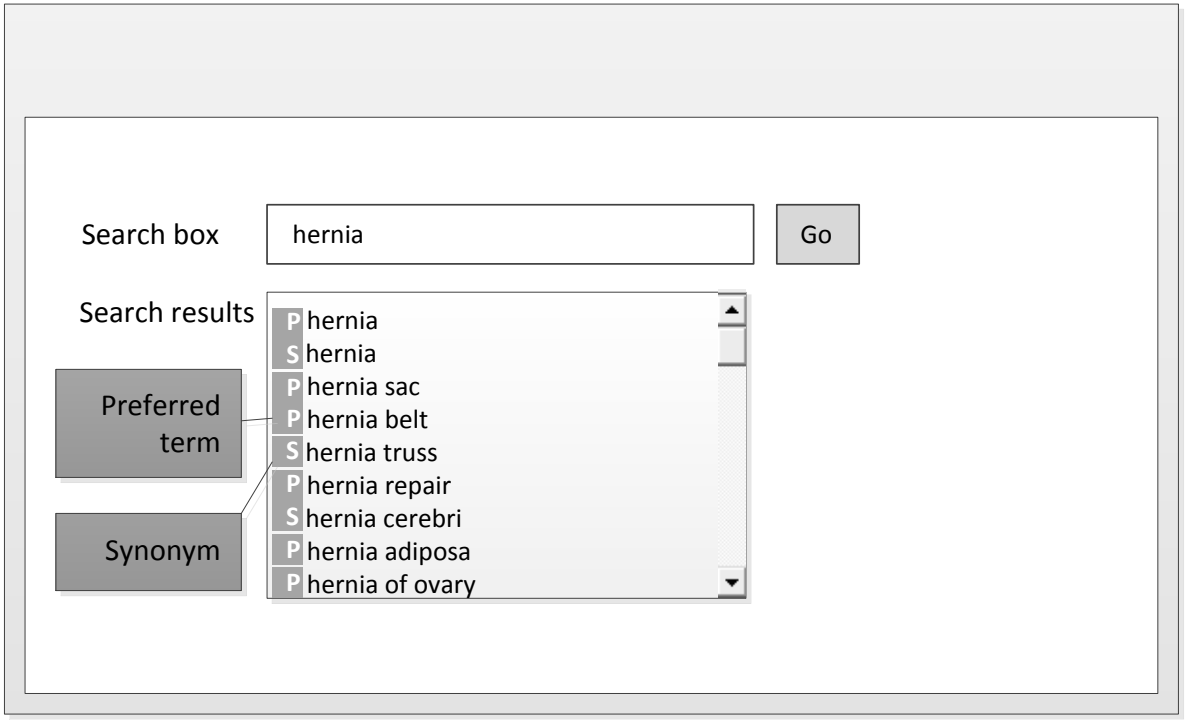
Order Search Results Rationally (1)

- Order shortest matching results first



Order Search Results Rationally (2)

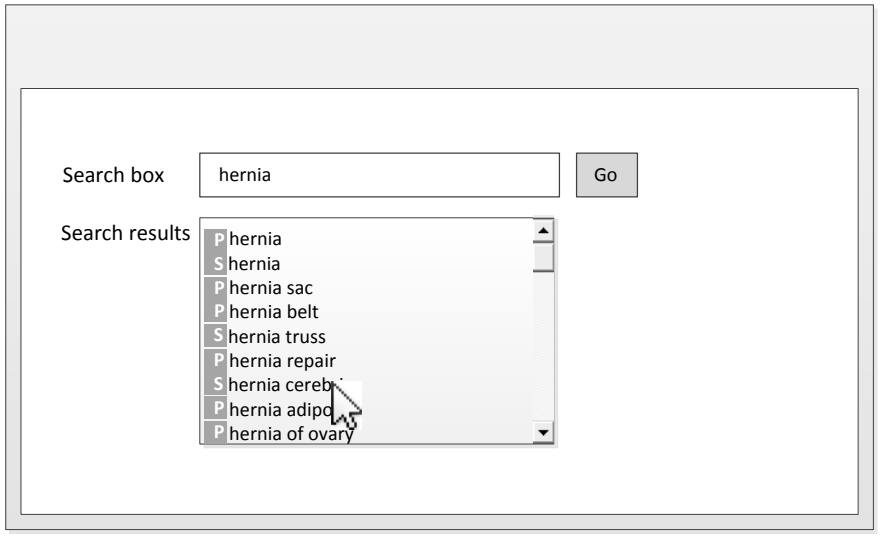
- Order preferred term matches before synonyms



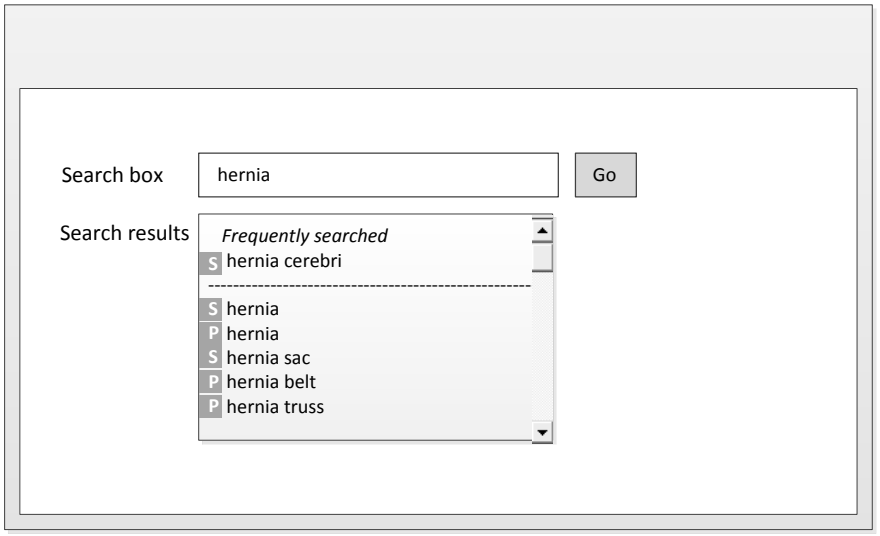
Order Search Results Rationally (5)

- Display search results with most frequently used descriptions listed first

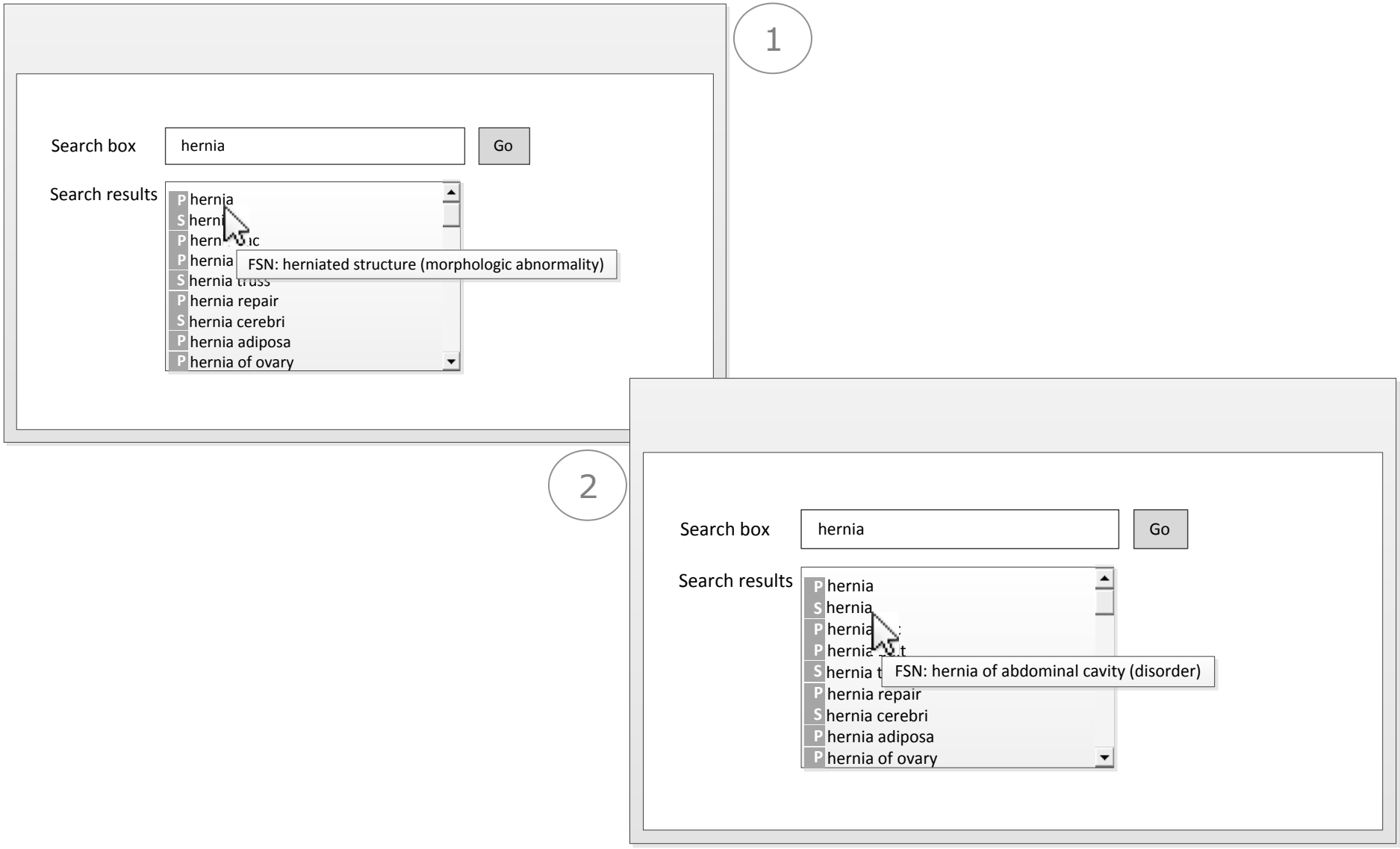
1



2

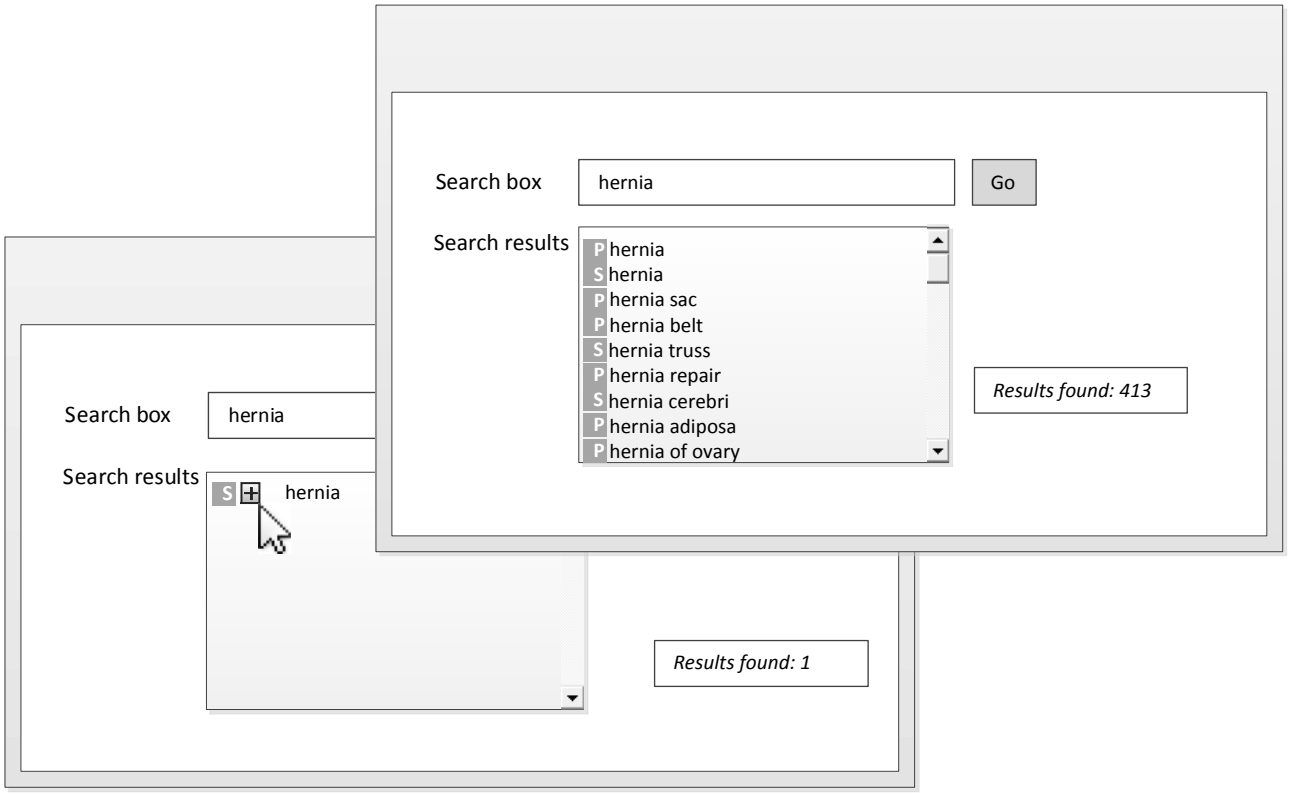


Distinguish identical descriptions of different concepts



Rationalize search results by subsumption checking

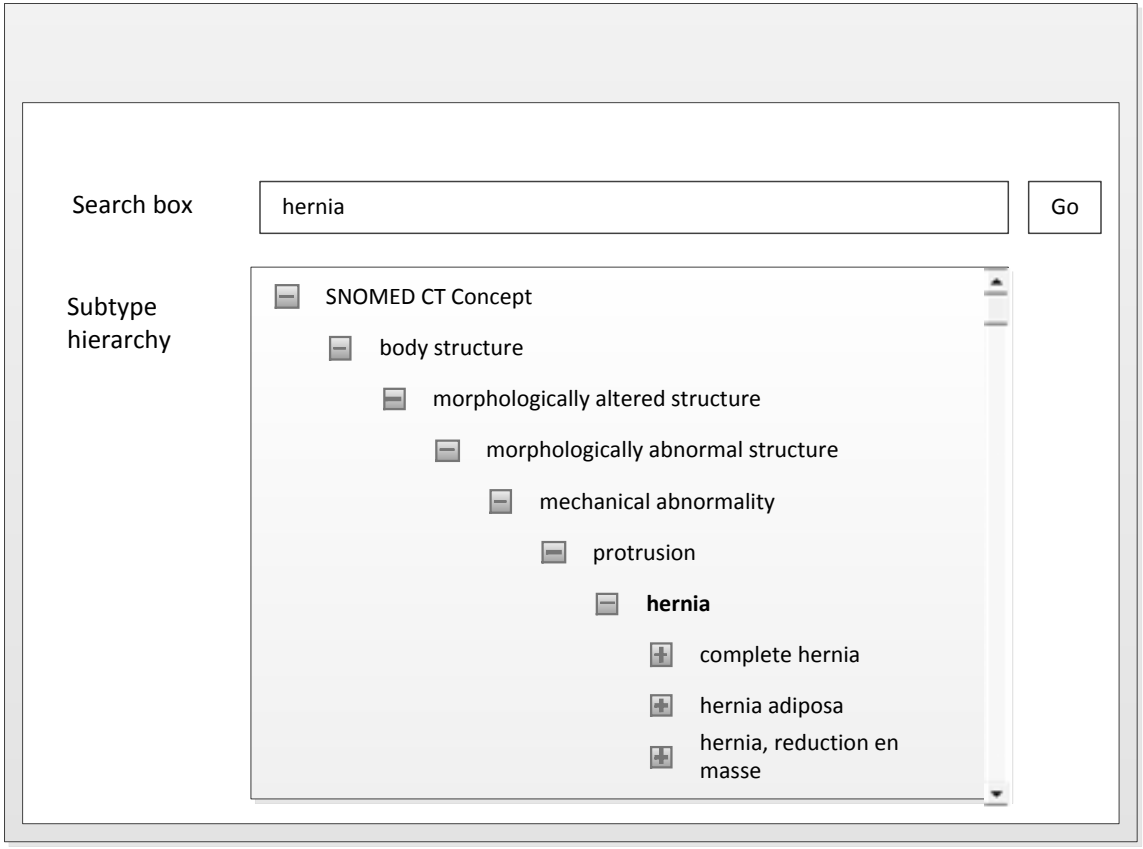
Before



After

Display navigation results effectively (1)

- Using the subtype hierarchy



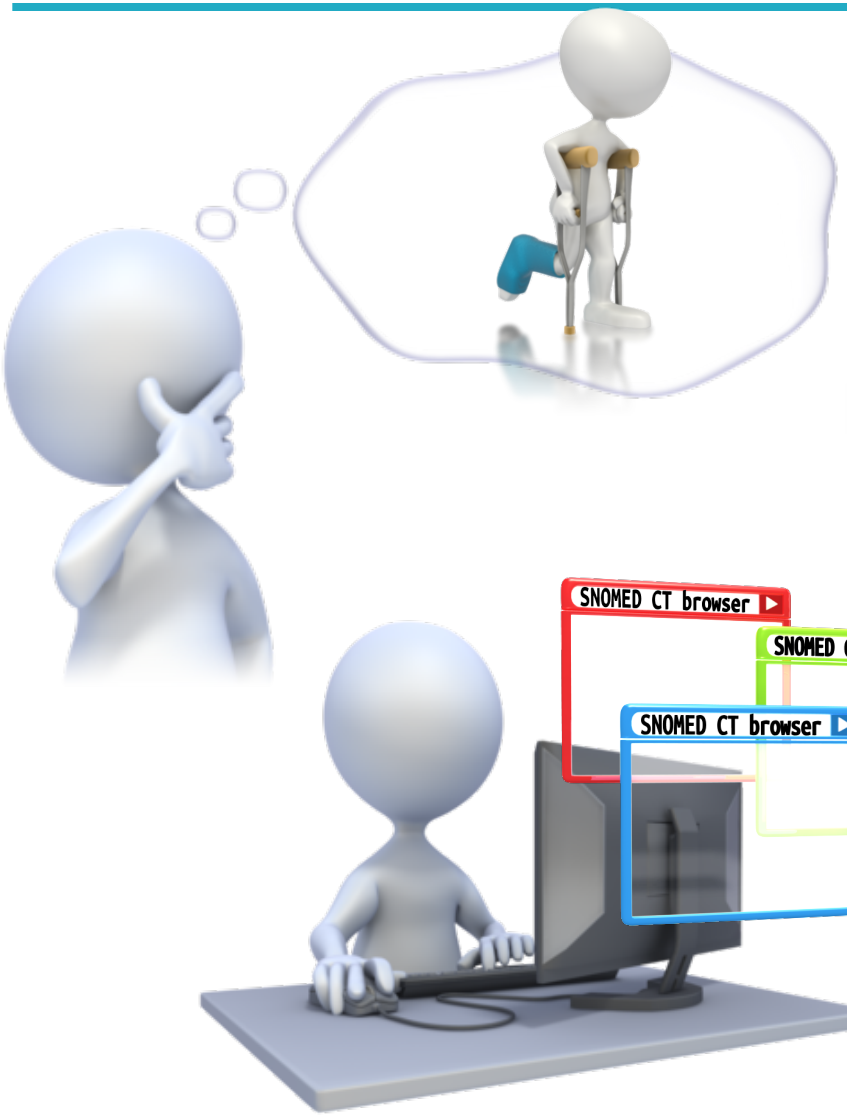
- Not designed for data entry

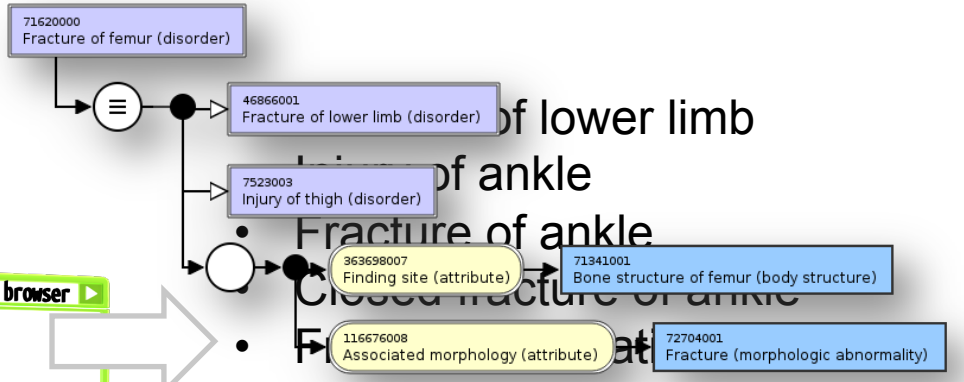
Display navigation results effectively (2)

- Using the navigation hierarchy (hand-crafted)

- Navigation hierarchies can be used to drive some types of structured data entry
- Navigation hierarchies can order data in sensible ways by priority, or by some readily understood convention (e.g. cranial nerve order).

Conclusive remarks





- Fracture of lower limb
- Injury of ankle
- Fracture of ankle
- Closed fracture of ankle
- Fracture of ankle joint
- Fracture of distal end of fibula
- Fracture of distal end of tibia
- Fracture of talus
- Open fracture of ankle

ihtsdo Leading healthcare terminology, worldwide

Thank you for your attention

Questions?

- Contact IHTSDO: info@ihtsdo.org
- Web site: www.ihtsdo.org

Delivering SNOMED CT The global language of healthcare

hscic Health & Social Care Information Centre

SNOMED CT

Search candidate generation

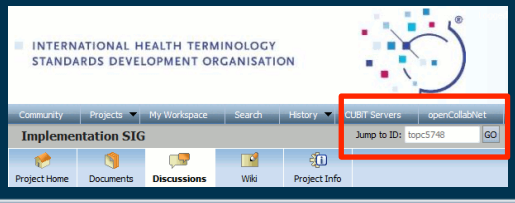
Dr Jeremy Rogers *jeremy.rogers@nhs.net*

IHTSDO Consultant Terminologist

Principal Terminology Specialists NHS HSCIC

IHTSDO Showcase, Amsterdam, October 30th 2014

See ImpSIG thread on Collabnet :
Jump to ID **topc5748** (page 2)



Take home point..

Lucene is good…

Lucene is a vector space search algorithm designed for full text indexing and searching of 'documents' containing paragraphs of text.

..but SNOMED CT isn't a newspaper.

'All descriptions on a concept' approximate poorly to the notion of 'a document'

'All descriptions in SNOMED' approximate very poorly to 'a collection of documents' from which Lucene may reasonably compute an inverse document frequency metric

Outline

Symptoms

Pathology

Treatment

Evidence

An Experiment:

16 clinical search phrases into 3 SNOMED CT browser implementations

pain lower back

low back pain

traveller diarrhoea

acute asthma

breast lump

breast mass

medication review

vaccine given

enlarged node

morning headache

staging scan

urge incontinence

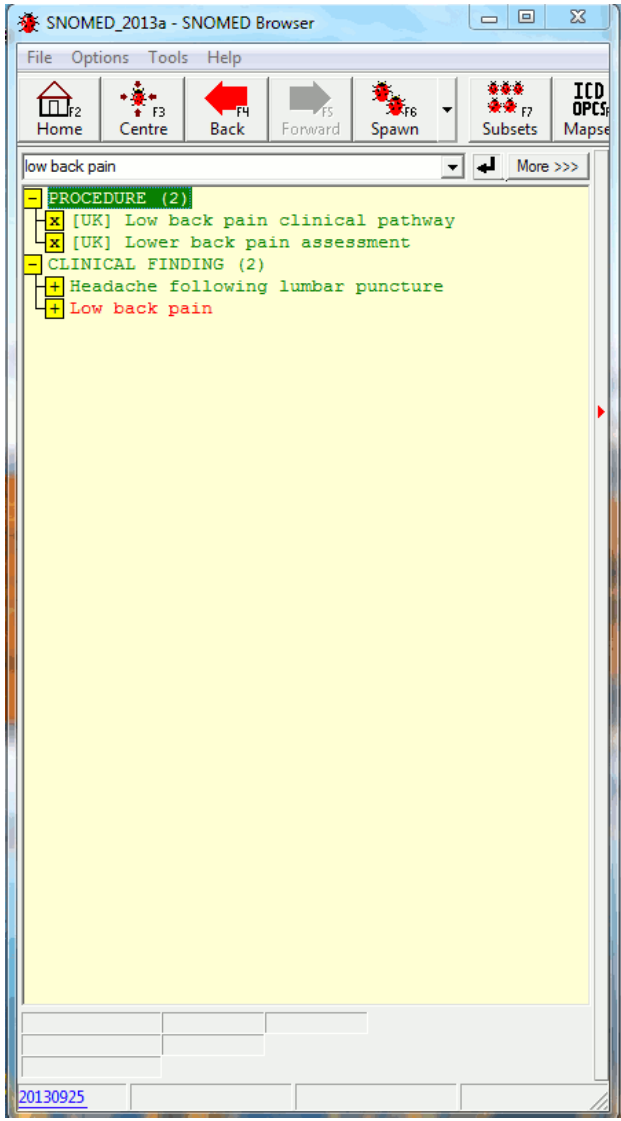
excision biopsy

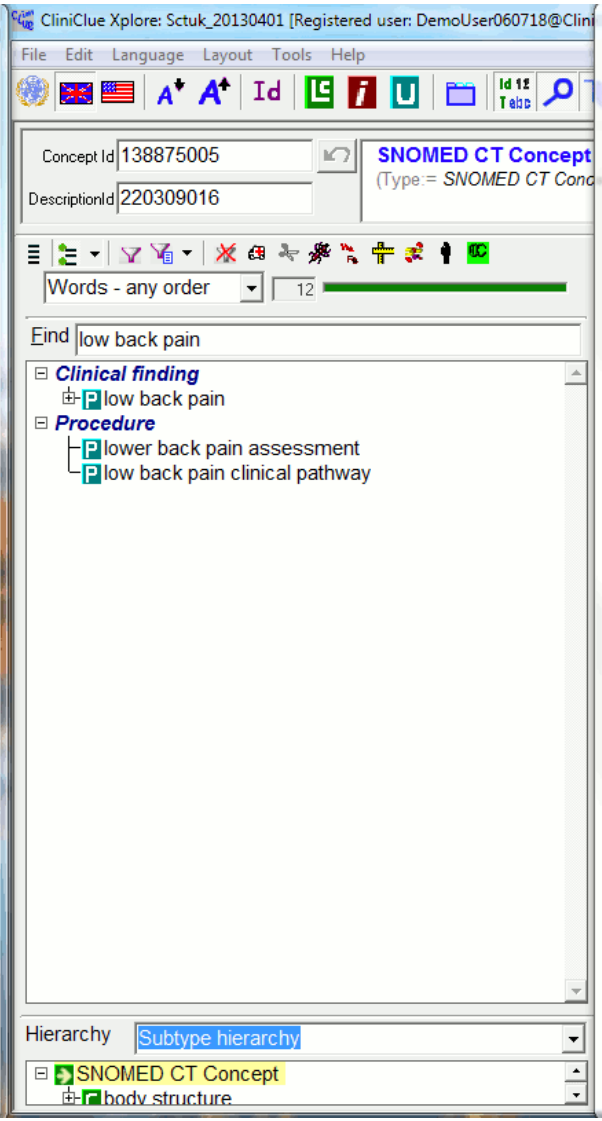
post operative analgesia

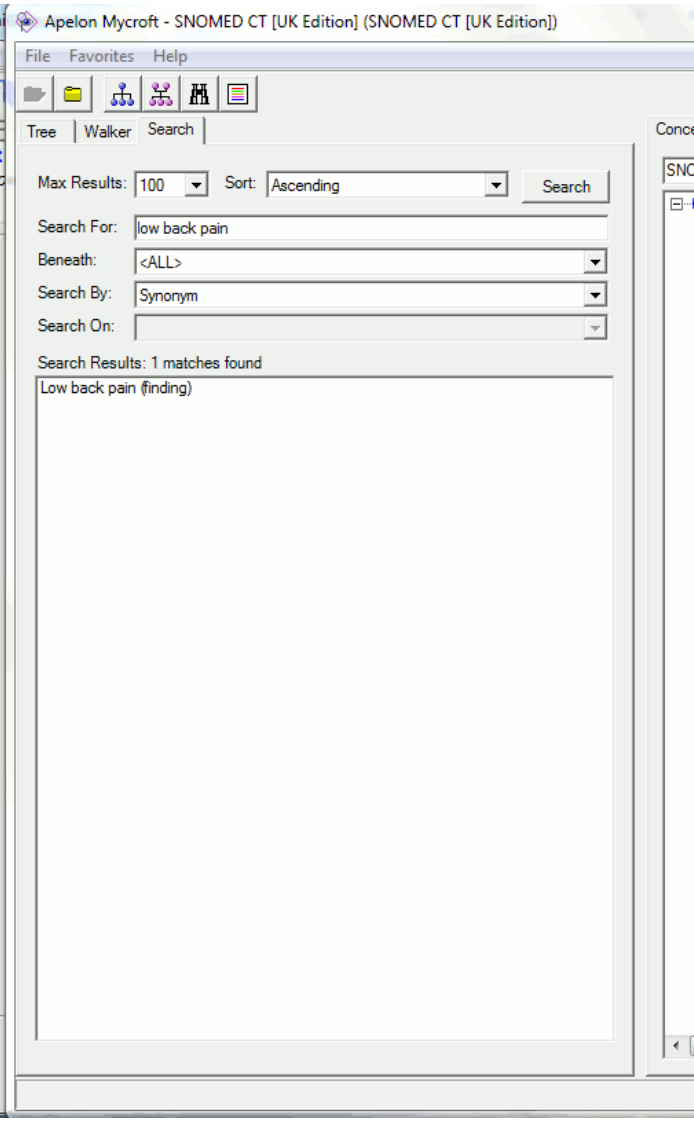
transoesophageal ultrasound

emergency caesarean

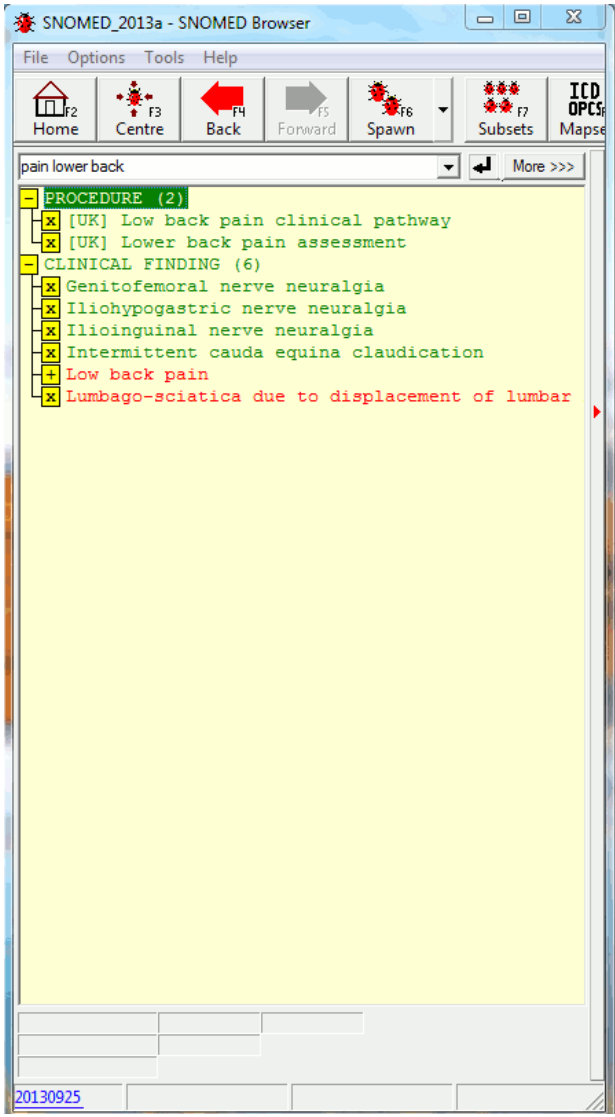
‘low back pain’

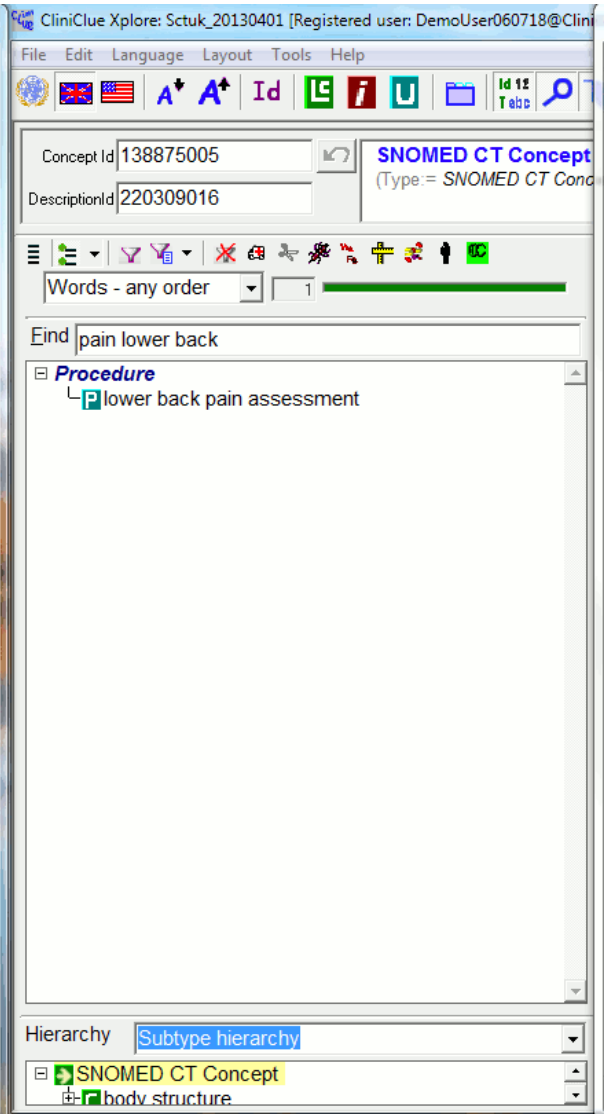


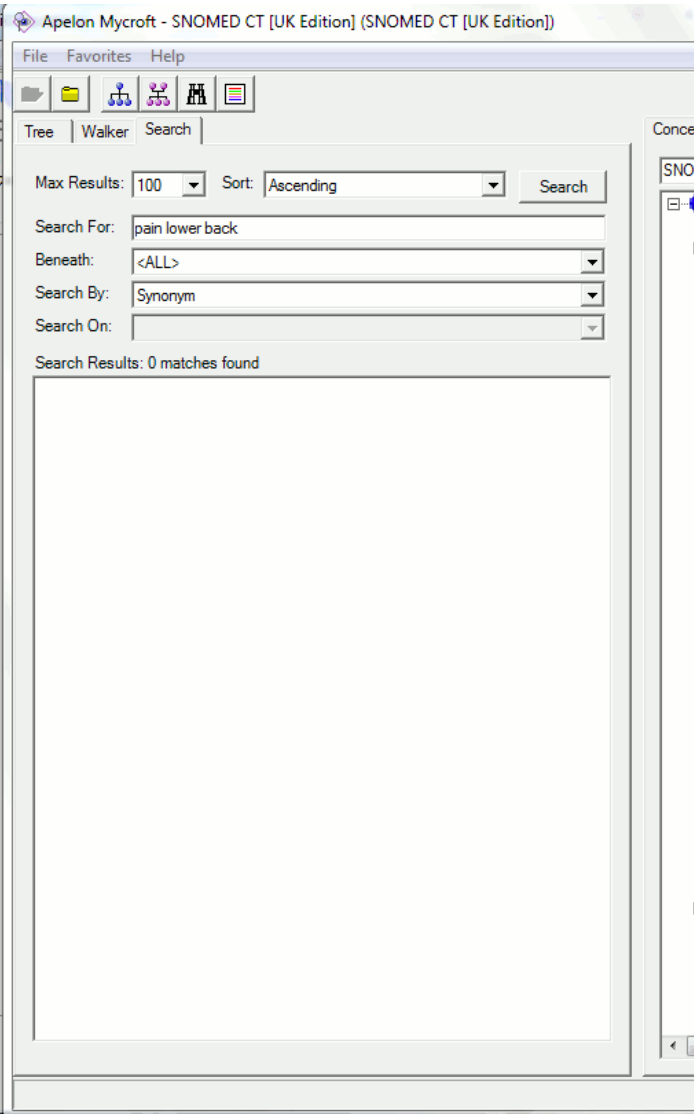




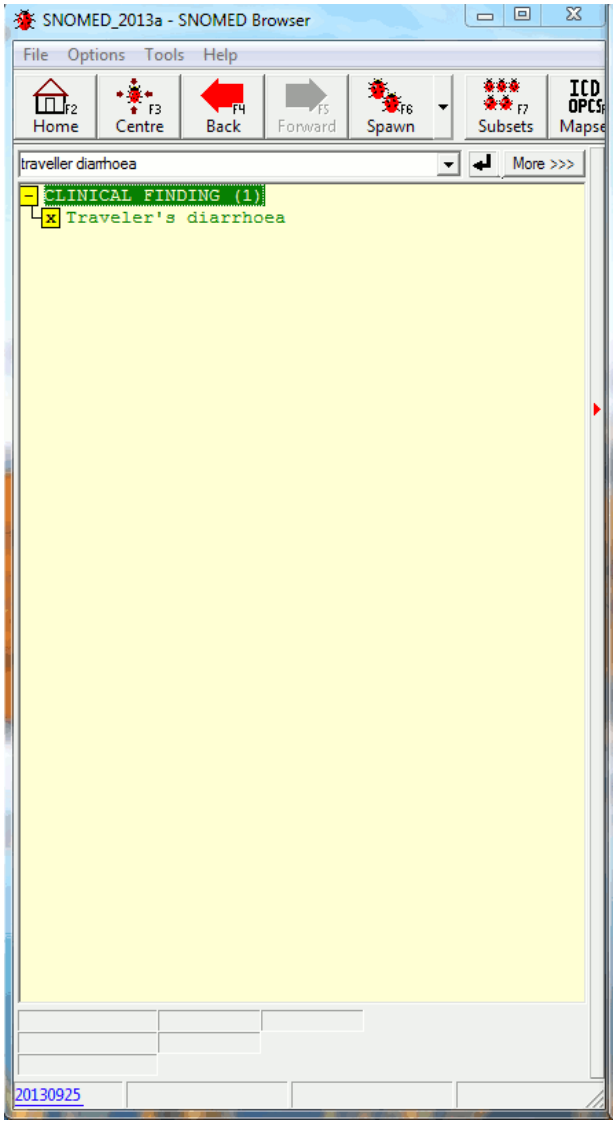
‘pain lower back’

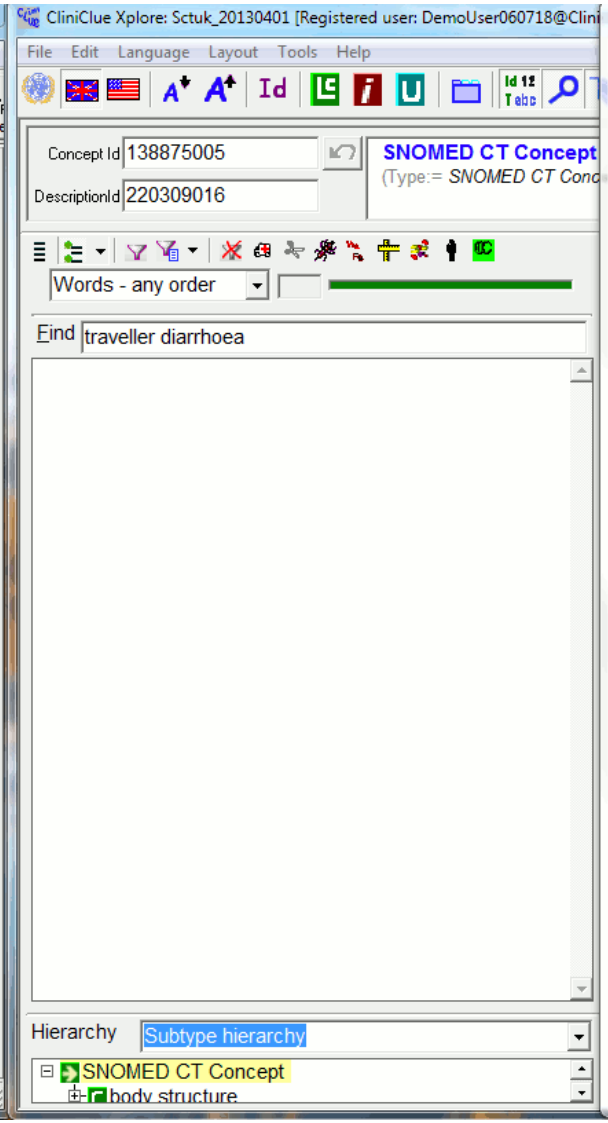


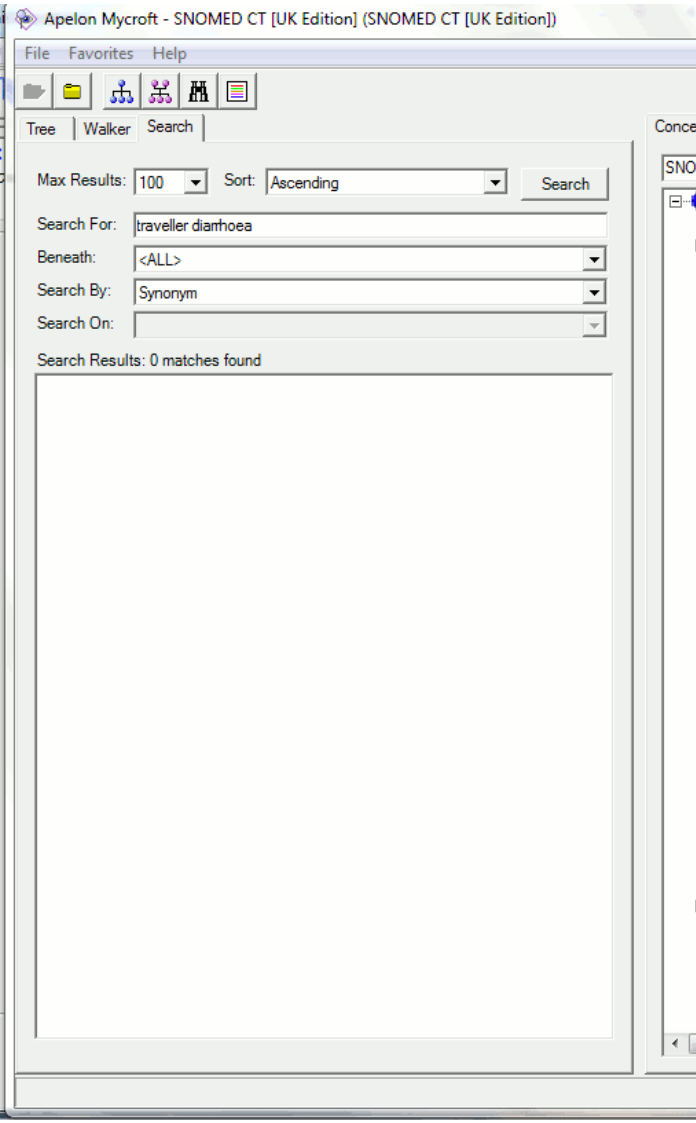




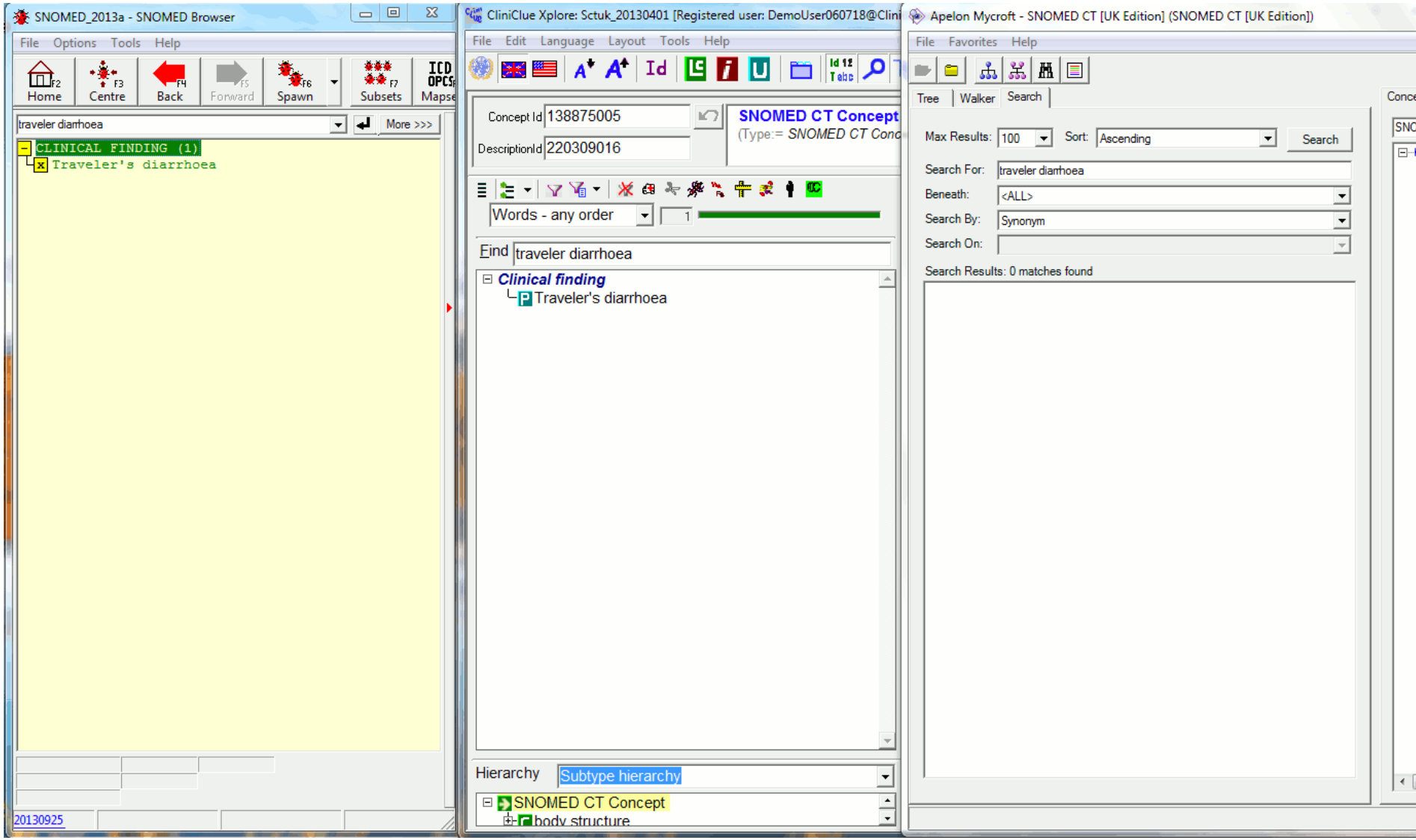
‘traveller diarrhoea’



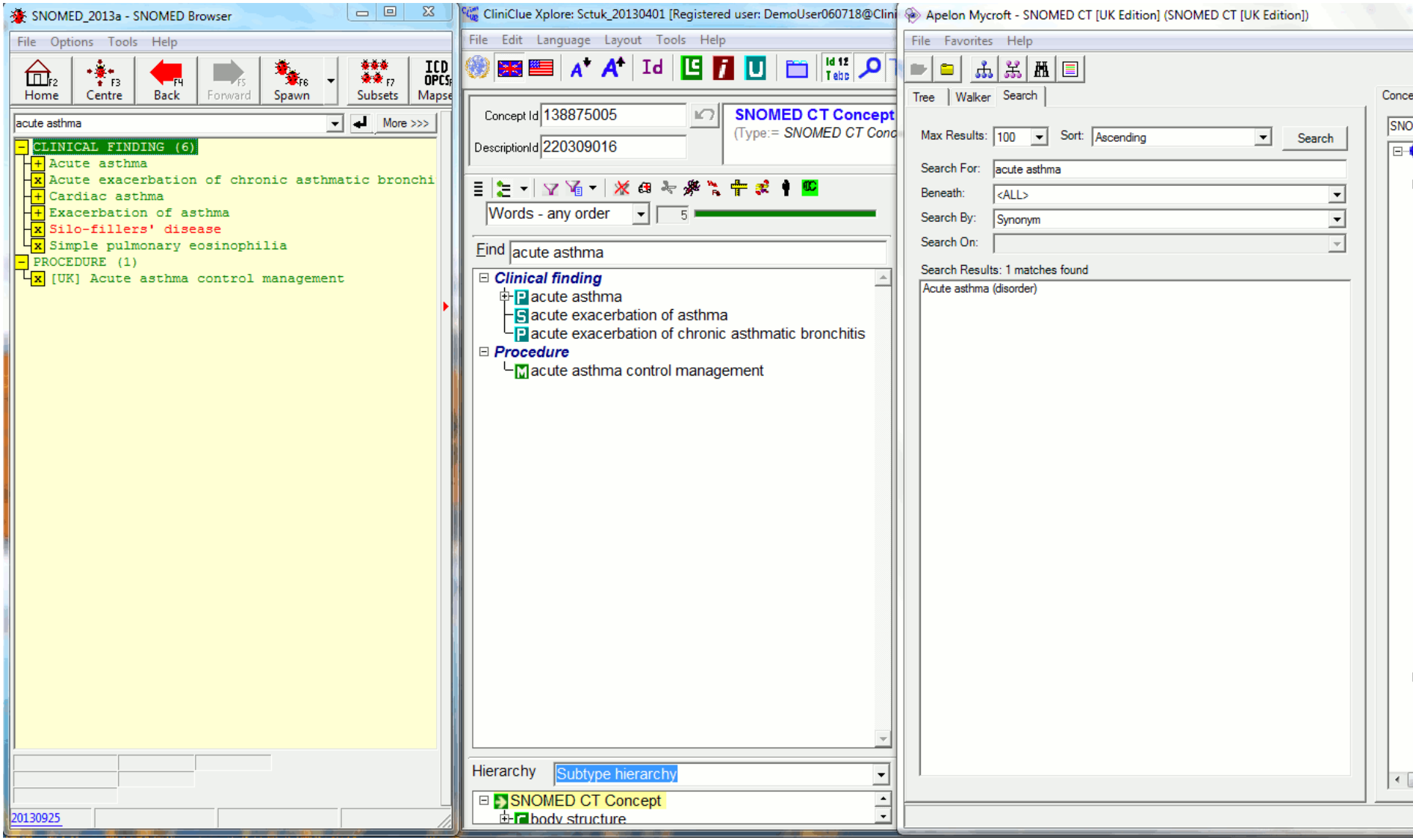




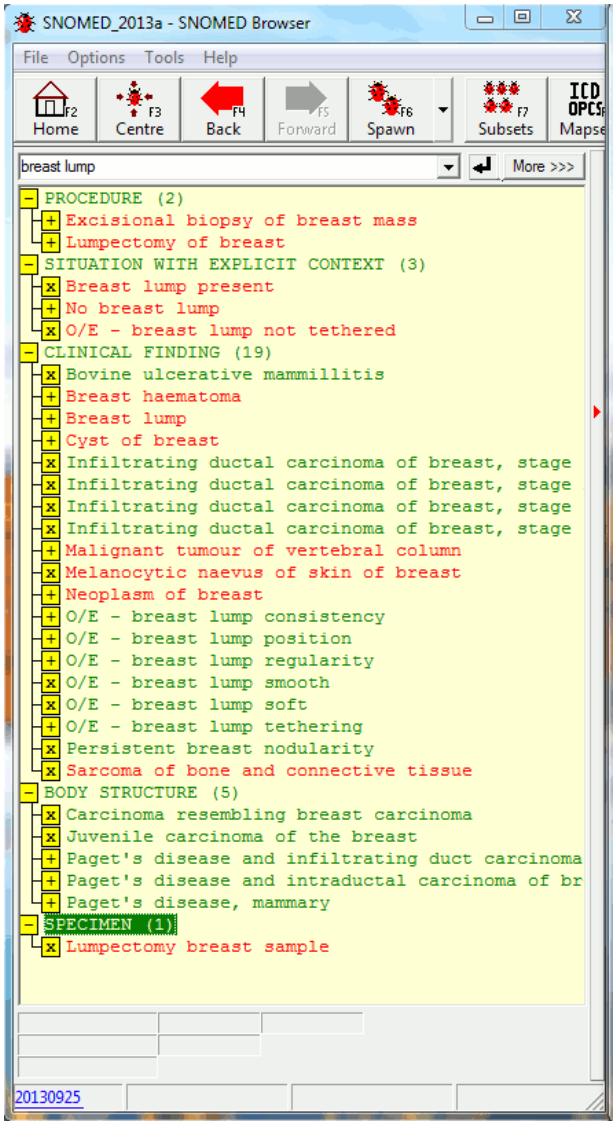
‘traveler diarrhoea

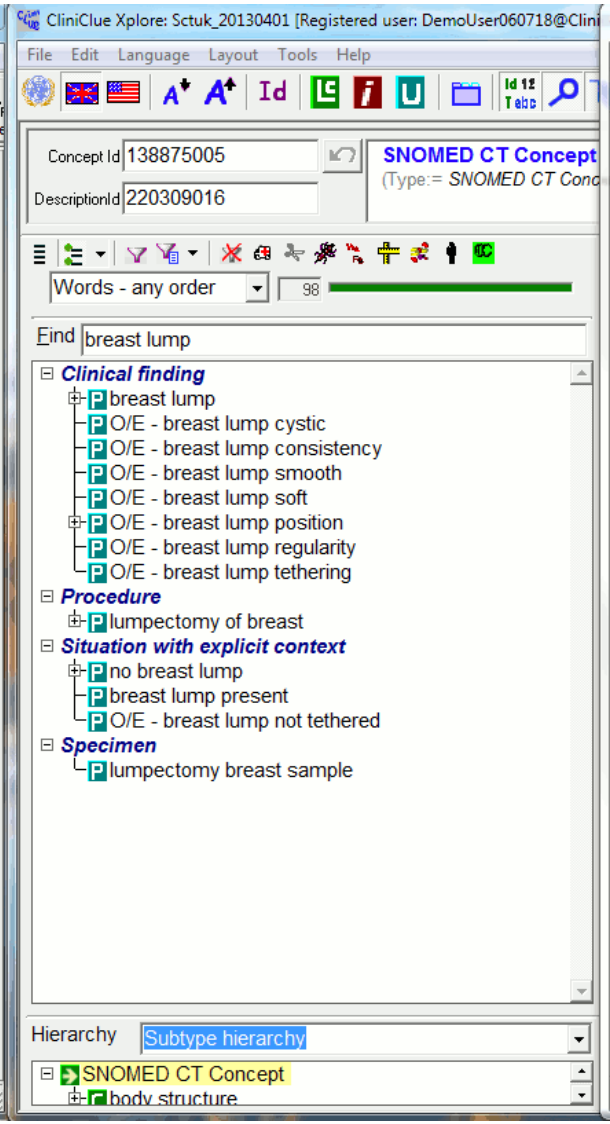


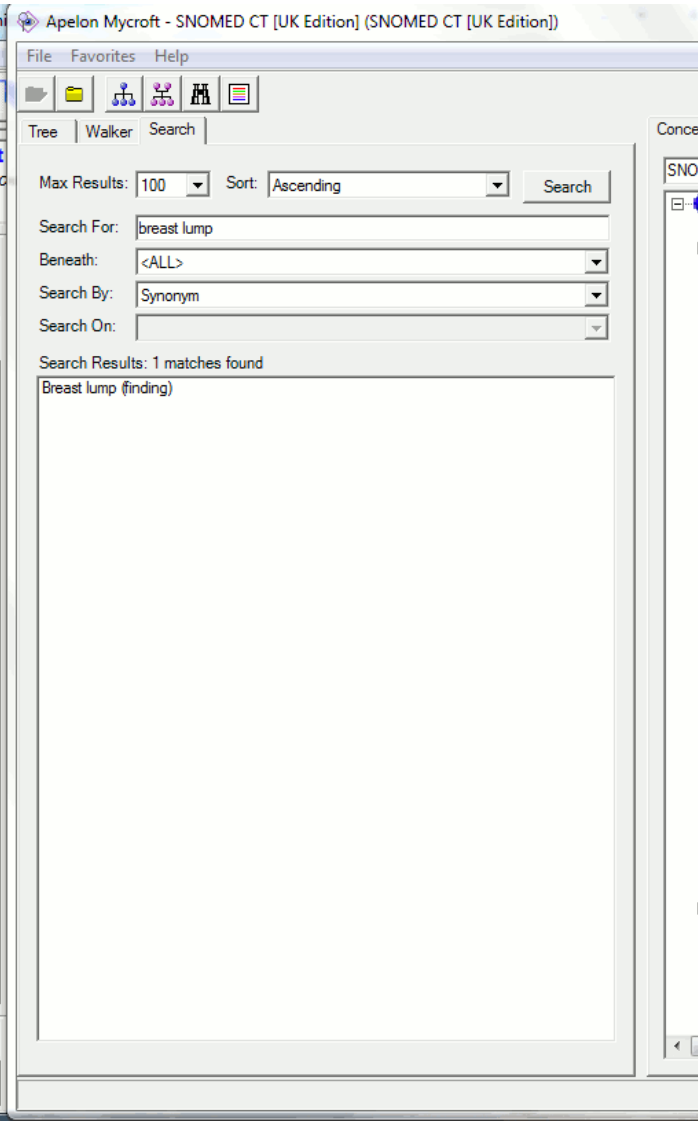
‘acute asthma’



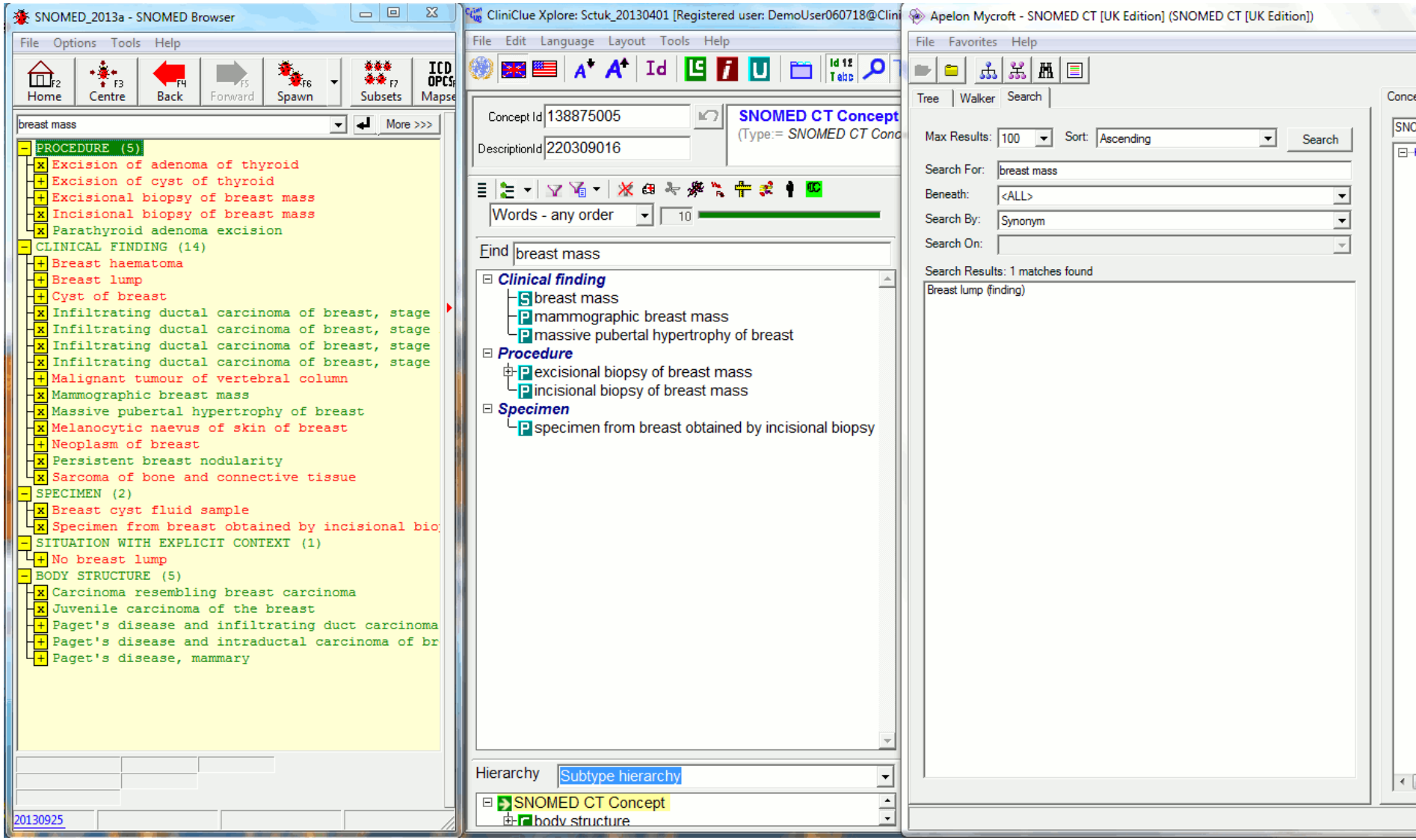
‘breast lump’


SNOMED_2013a - SNOMED Browser
File Options Tools Help
Home Centre Back Forward Spawn Subsets Mapse
breast lump
More >>>
PROCEDURE (2)
Excisional biopsy of breast mass
Lumpectomy of breast
SITUATION WITH EXPLICIT CONTEXT (3)
Breast lump present
No breast lump
O/E - breast lump not tethered
CLINICAL FINDING (19)
Bovine ulcerative mammillitis
Breast haematoma
Breast lump
Cyst of breast
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Malignant tumour of vertebral column
Melanocytic naevus of skin of breast
Neoplasm of breast
O/E - breast lump consistency
O/E - breast lump position
O/E - breast lump regularity
O/E - breast lump smooth
O/E - breast lump soft
O/E - breast lump tethering
Persistent breast nodularity
Sarcoma of bone and connective tissue
BODY STRUCTURE (5)
Carcinoma resembling breast carcinoma
Juvenile carcinoma of the breast
Paget's disease and infiltrating duct carcinoma
Paget's disease and intraductal carcinoma of br
Paget's disease, mammary
SPECIMEN (1)
Lumpectomy breast sample
20130925

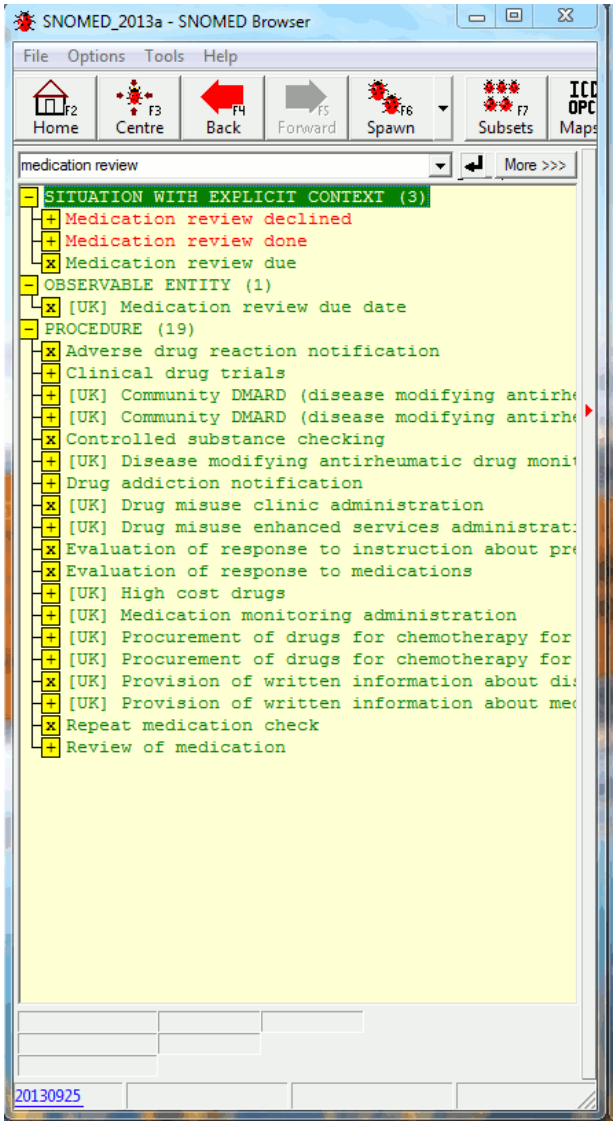

CliniClue Xplore: Sctuk_20130401 [Registered user: DemoUser060718@Clini
File Edit Language Layout Tools Help
Concept Id 138875005
DescriptionId 220309016
SNOMED CT Concept
(Type:= SNOMED CT Conc
Words - any order
98
Find breast lump
Clinical finding
breast lump
O/E - breast lump cystic
O/E - breast lump consistency
O/E - breast lump smooth
O/E - breast lump soft
O/E - breast lump position
O/E - breast lump regularity
O/E - breast lump tethering
Procedure
lumpectomy of breast
Situation with explicit context
no breast lump
breast lump present
O/E - breast lump not tethered
Specimen
lumpectomy breast sample
Hierarchy
Subtype hierarchy
SNOMED CT Concept
body structure

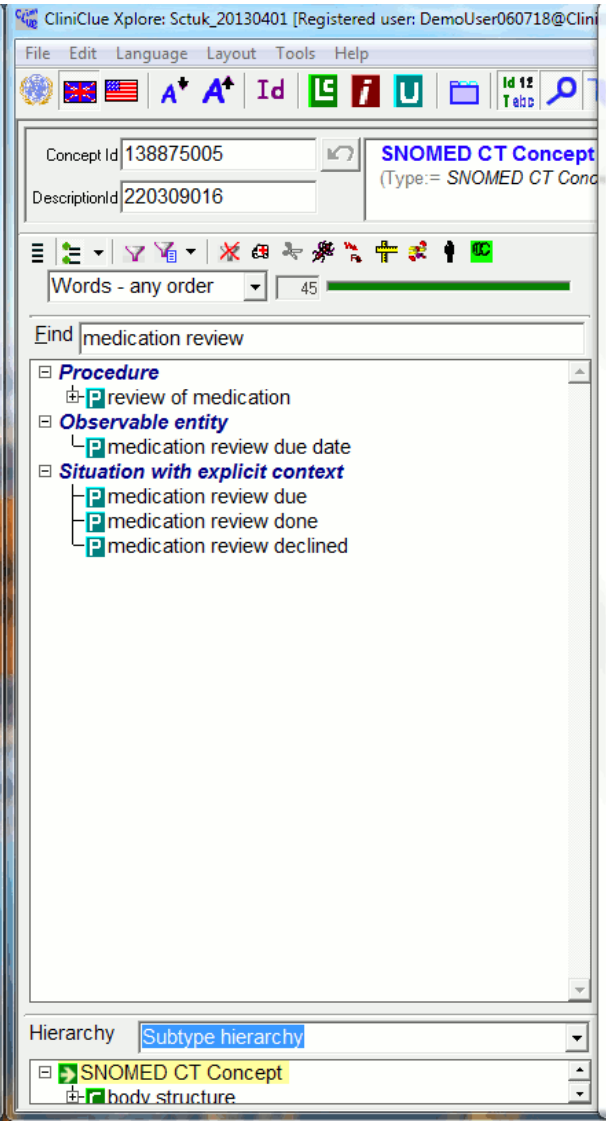

Apelon Mycroft - SNOMED CT [UK Edition] (SNOMED CT [UK Edition])
File Favorites Help
Tree Walker Search
Max Results: 100
Sort: Ascending
Search
Search For: breast lump
Beneath: <ALL>
Search By: Synonym
Search On:
Search Results: 1 matches found
Breast lump (finding)

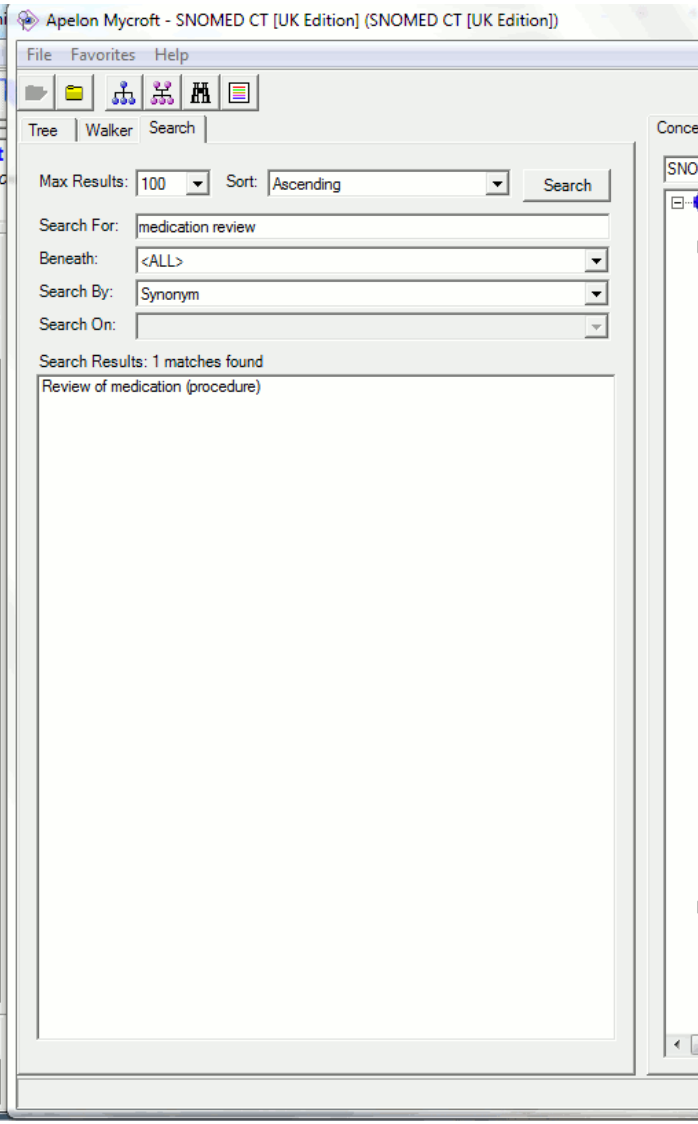
‘breast *mass*’


SNOMED_2013a - SNOMED Browser
File Options Tools Help
Home Centre Back Forward Spawn Subsets Mapse
breast mass
More >>>
PROCEDURE (5)
Excision of adenoma of thyroid
Excision of cyst of thyroid
Excisional biopsy of breast mass
Incisional biopsy of breast mass
Parathyroid adenoma excision
CLINICAL FINDING (14)
Breast haematoma
Breast lump
Cyst of breast
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Infiltrating ductal carcinoma of breast, stage
Malignant tumour of vertebral column
Mammographic breast mass
Massive pubertal hypertrophy of breast
Melanocytic naevus of skin of breast
Neoplasm of breast
Persistent breast nodularity
Sarcoma of bone and connective tissue
SPECIMEN (2)
Breast cyst fluid sample
Specimen from breast obtained by incisional bio
SITUATION WITH EXPLICIT CONTEXT (1)
No breast lump
BODY STRUCTURE (5)
Carcinoma resembling breast carcinoma
Juvenile carcinoma of the breast
Paget's disease and infiltrating duct carcinoma
Paget's disease and intraductal carcinoma of br
Paget's disease, mammary
20130925
CliniClue Xplore: Sctuk_20130401 [Registered user: DemoUser060718@Clini
File Edit Language Layout Tools Help
Concept Id 138875005
DescriptionId 220309016
SNOMED CT Concept
(Type:= SNOMED CT Conc
Words - any order
10
Find breast mass
Clinical finding
breast mass
mammographic breast mass
massive pubertal hypertrophy of breast
Procedure
excisional biopsy of breast mass
incisional biopsy of breast mass
Specimen
specimen from breast obtained by incisional biopsy
Hierarchy
Subtype hierarchy
SNOMED CT Concept
body structure
Apelon Mycroft - SNOMED CT [UK Edition] (SNOMED CT [UK Edition])
File Favorites Help
Tree Walker Search
Max Results: 100
Sort: Ascending
Search
Search For: breast mass
Beneath: <ALL>
Search By: Synonym
Search On:
Search Results: 1 matches found
Breast lump (finding)

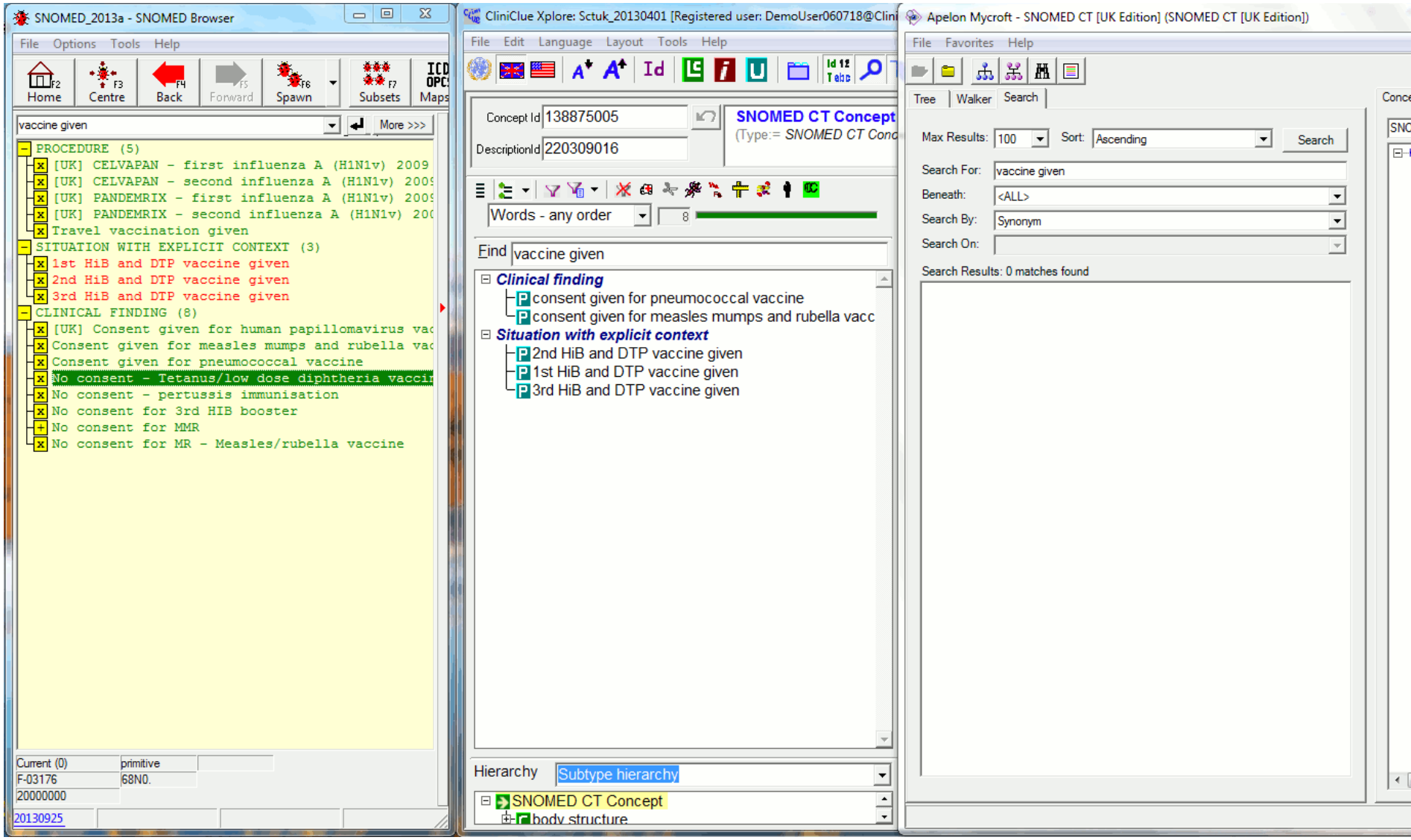
‘medication review’







‘vaccine given’



An Experiment:

16 clinical search phrases into 3 SNOMED CT browser implementations

pain lower back	enlarged node	4+	5+	0
low back pain	morning headache	1	1	1
traveller diarrhoea	staging scan	7	0	0
acute asthma	urge incontinence	3	3	1
breast lump	excision biopsy	45+	22+	1
breast mass	post operative analgesia	1	0	0
medication review	transoesophageal ultrasound	2	1	0
vaccine given	emergency caesarean	3	3	0

SNOMED CT Miscodes

39 months in a busy UK A&E Department

- Setting: 408,831 coded ED episodes
  - One ‘reason’ code per completed visit
  - 39 months (Oct 2008 – Dec 2011)
  - 12,022 distinct SCT codes selected at least once
- Users: No training, time pressured
- Browser: term match on all of SNOMED CT hierarchy not shown
- Result: 11% of data is seriously miscoded

SNOMED CT: New ways to get it wrong

‘Obvious’ miscode examples

1097	Temperature	246508008|Temperature (attribute)|
145	High temperature	285717004|High temperature (physical force)|
118	FB (Foreign Body)	367409002|Followed by (attribute)|
24	TB (Tuberculosis)	60117003|Transmitted by (attribute)|
33	Spot	23840004|Leiostomus xanthurus (organism)|
15	MI (Myocardial Infarct)	45169001|Without (attribute)|
17	Drug used	246488008|Drug used (attribute)|
8	Drugs	2280110000000101|Drugs (navigational concept)|
373	ETOH - Alcohol intake	160573003|Alcohol intake (observable entity)|
207	Alcohol	53041004|Alcohol (substance)|
117	EtOH – Ethanol	419442005|Ethyl alcohol (substance)|
50	Ethanol	419442005|Ethyl alcohol (substance)|
33	Lymph node	59441001|Structure of lymph node (body structure)|
136	Nasogastric tube	17102003|Nasogastric tube, device (physical object)|
82	Catheter	19923001|Catheter, device (physical object)|
535	RTA	1776003|Renal tubular acidosis (disorder)|
396	Psychiatric	27296002|Psychiatric (qualifier value)|
230	Stabbing	410706007|Stabbing sensation quality (qualifier value)|

Outline

Symptoms

Pathology

Treatment

Evidence

The Problem: CodeForString(str) operation

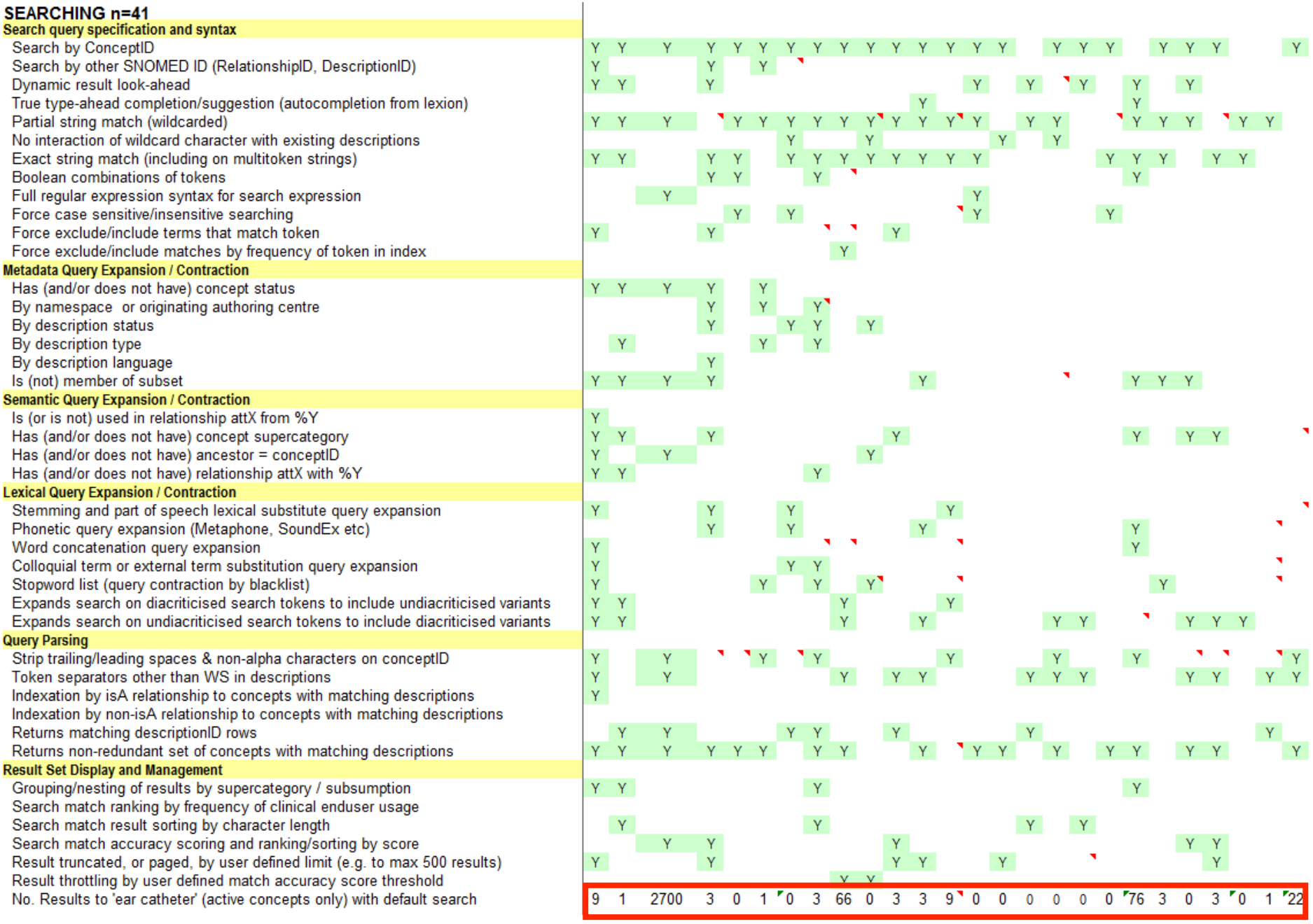
Many different strategies can be employed

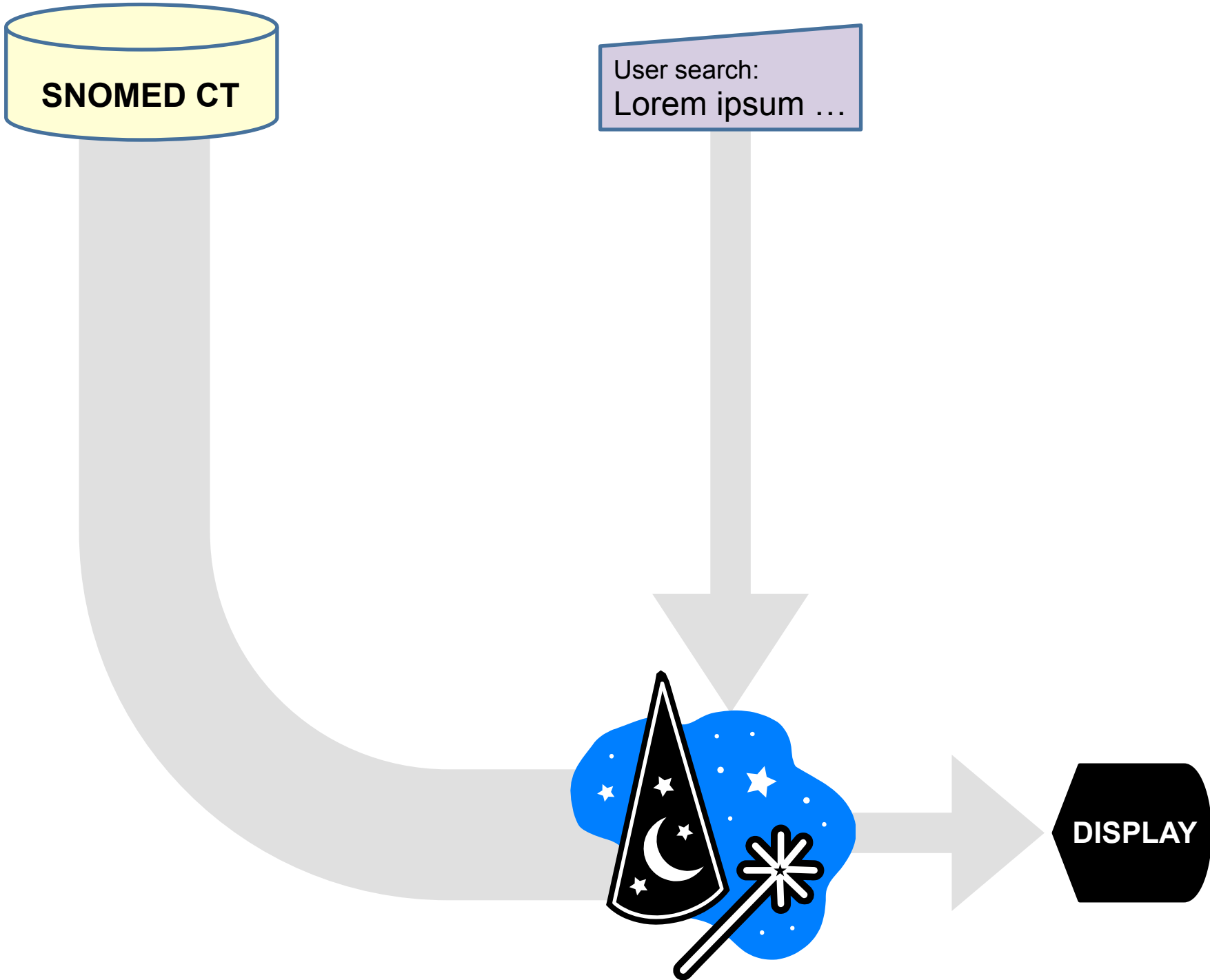
- All produce different false +ve and –ve results
- At least one existing implementation for most elements
- Very few implementations with all elements
- Naïve approaches predominate
- But clever strategies can be slow
  - and need extra lexical resources

Different strategies ⇨ different results

Different results ⇨ interrater variability

Interrater variability ⇨ reduced interoperability


SEARCHING n=41
Search query specification and syntax
Search by ConceptID
Search by other SNOMED ID (RelationshipID, DescriptionID)
Dynamic result look-ahead
True type-ahead completion/suggestion (autocompletion from lexion)
Partial string match (wildcarded)
No interaction of wildcard character with existing descriptions
Exact string match (including on multitoken strings)
Boolean combinations of tokens
Full regular expression syntax for search expression
Force case sensitive/insensitive searching
Force exclude/include terms that match token
Force exclude/include matches by frequency of token in index
Metadata Query Expansion / Contraction
Has (and/or does not have) concept status
By namespace or originating authoring centre
By description status
By description type
By description language
Is (not) member of subset
Semantic Query Expansion / Contraction
Is (or is not) used in relationship attX from %Y
Has (and/or does not have) concept supercategory
Has (and/or does not have) ancestor = conceptID
Has (and/or does not have) relationship attX with %Y
Lexical Query Expansion / Contraction
Stemming and part of speech lexical substitute query expansion
Phonetic query expansion (Metaphone, SoundEx etc)
Word concatenation query expansion
Colloquial term or external term substitution query expansion
Stopword list (query contraction by blacklist)
Expands search on diacriticised search tokens to include undiacriticised variants
Expands search on undiacriticised search tokens to include diacriticised variants
Query Parsing
Strip trailing/leading spaces & non-alpha characters on conceptID
Token separators other than WS in descriptions
Indexation by isA relationship to concepts with matching descriptions
Indexation by non-isA relationship to concepts with matching descriptions
Returns matching descriptionID rows
Returns non-redundant set of concepts with matching descriptions
Result Set Display and Management
Grouping/nesting of results by supercategory / subsumption
Search match ranking by frequency of clinical enduser usage
Search match result sorting by character length
Search match accuracy scoring and ranking/sorting by score
Result truncated, or paged, by user defined limit (e.g. to max 500 results)
Result throttling by user defined match accuracy score threshold
No. Results to 'ear catheter' (active concepts only) with default search
9 1 2700 3 0 1 0 3 66 0 3 3 9 0 0 0 0 0 0 76 3 0 3 0 1 22


SNOMED CT
User search:
Lorem ipsum …
DISPLAY

Outline

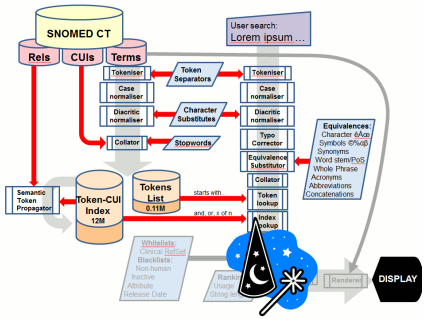
Symptoms

Pathology

Treatment

Evidence

From term to concept…



TOKENISATION

- Non WS token separators
- Character stripping

TOKEN SET EXPANSION

- Word equivalence[TIG 6.5.4]
- Diacritic/Char substn
- Metaphone / SoundEx
- Levenshtein Distance
- Token concatenation

TOKEN CONTRACTION

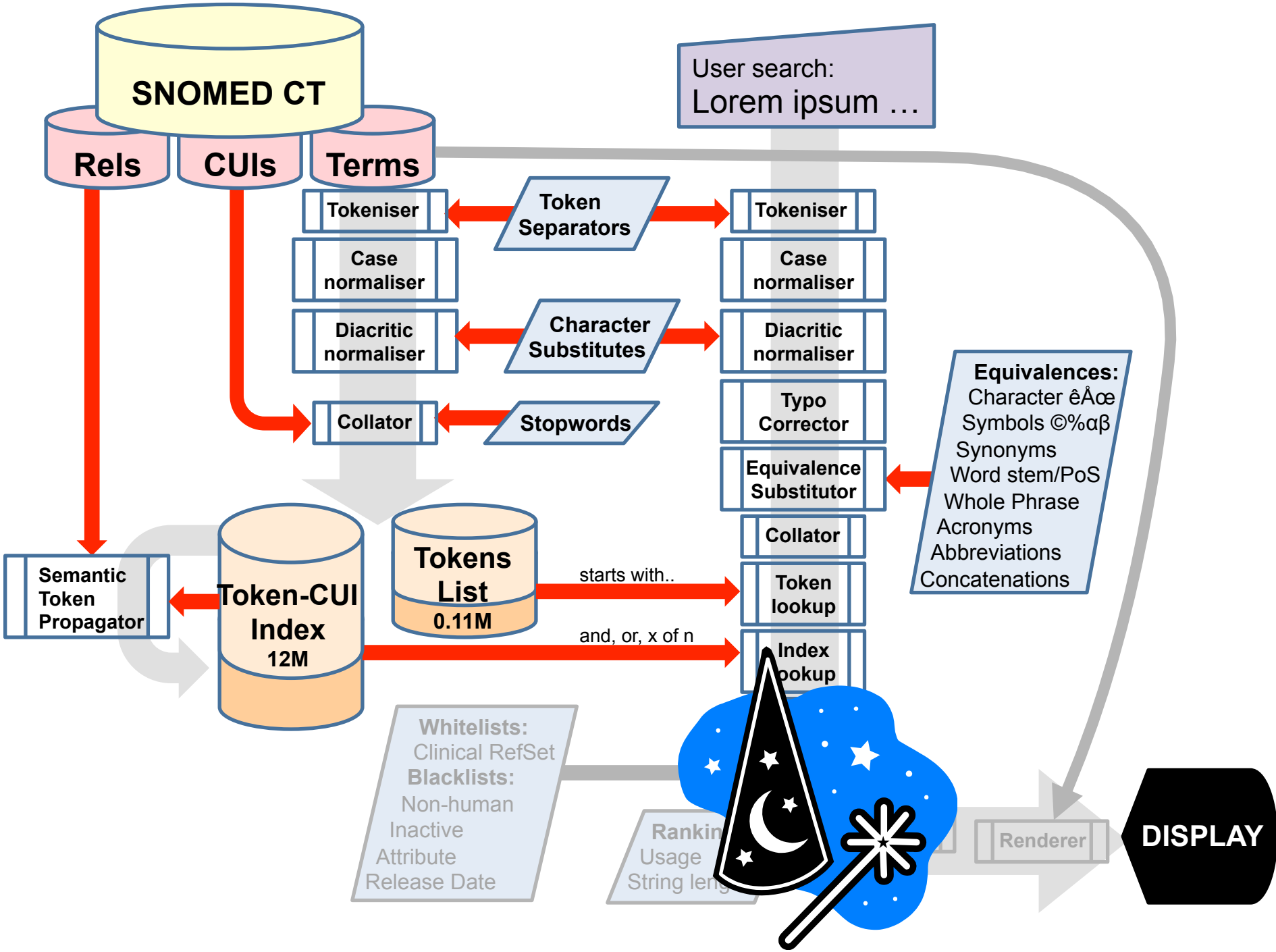
- Stopword lists
- Word stemming

TOKEN LOOKUP

- Wildcard characters in corpus
- Exact matching
- Case sensitivity
- Force include/exclude tokens
- Query token proximity
- Query token (re-)ordering[TIG]
- Default wildcarding
- Multitoken semantics[TIG]
- Token inheritance

Match scoring and ranking

Match grouping & ordering


SNOMED CT
Rels
CUIs
Terms
User search:
Lorem ipsum …
Tokeniser
Token Separators
Tokeniser
Case normaliser
Case normaliser
Diacritic normaliser
Character Substitutes
Diacritic normaliser
Collator
Stopwords
Typo Corrector
Equivalence Substitutor
Equivalences:
Character êÅœ
Symbols ©%αβ
Synonyms
Word stem/PoS
Whole Phrase
Acronyms
Abbreviations
Concatenations
Collator
Semantic Token Propagator
Token-CUI Index
12M
Tokens List
0.11M
starts with..
Token lookup
and, or, x of n
Index lookup
Whitelists:
Clinical RefSet
Blacklists:
Non-human
Inactive
Attribute
Release Date
Usage
Renderer
DISPLAY

IHTSDO Lexical Resources Stopwords

- zres_ExcludedWords_en-US_INT_20070731.txt
- (n=53) Virtually unchanged since 2002

IHTSDO Stopwords 2013			(2002 version)
ABOUT	CANNOT	MORE	SPECIFIC
ALONGSIDE	CHRONICALLY	MUST	THAN
AN	CONSISTS	NO	THAT
AND	COVERED	NOT	THE
ANYTHING	DOES	OF	THINGS
AROUND	DURING	ON	THIS
AS	EVERY	ONLY	THROUGHOUT
AT	FINDS	OR	TO
BECAUSE	FOR	PROPERLY	UP
BEFORE	FROM	SIDE	USING
BEING	IN	SIDED	USUALLY
BOTH	INSTEAD	SOME	WHEN
BY	INTO	SOMETHING	WHILE
			WITHOUT

DOWN?

WITH?

mySQL 5.5 Stopword List

mySQL Full Text Search Stopword List (2011)

a's	able	about	above	according	her	here	here's	hereafter	hereby	serious	seriously	seven	several	shall
accordingly	across	actually	after	afterwards	herein	hereupon	hers	herself	hi	she	should	shouldn't	since	six
again	against	ain't	all	allow	him	himself	his	hither	hopefully	so	some	somebody	somehow	someone
allows	almost	alone	along	already	how	howbeit	however	i'd	i'll	something	sometime	sometimes	somewhat	somewhere
also	although	always	am	among	i'm	i've	ie	if	ignored	soon	sorry	specified	specify	specifying
amongst	an	and	another	any	immediate	in	inasmuch	inc	indeed	still	sub	such	sup	sure
anybody	anyhow	anyone	anything	anyway	indicate	indicated	indicates	inner	insofar	t's	take	taken	tell	tends
anyways	anywhere	apart	appear	appreciate	instead	into	inward	is	isn't	th	than	thank	thanks	thanx
appropriate	are	aren't	around	as	it	it'd	it'll	it's	its	that	that's	thats	the	their
aside	ask	asking	associated	at	itself	just	keep	keeps	kept	theirs	them	themselves	then	thence
available	away	awfully	be	became	know	known	knows	last	lately	there	there's	thereafter	thereby	therefore
because	become	becomes	becoming	been	later	latter	latterly	least	less	therein	theres	thereupon	these	they
before	beforehand	behind	being	believe	lest	let	let's	like	liked	they'd	they'll	they're	they've	think
below	beside	besides	best	better	likely	little	look	looking	looks	third	this	thorough	thoroughly	those
between	beyond	both	brief	but	ltd	mainly	many	may	maybe	though	three	through	throughout	thru
by	c'mon	c's	came	can	me	mean	meanwhile	merely	might	thus	to	together	too	took
can't	cannot	cant	cause	causes	more	moreover	most	mostly	much	toward	towards	tried	tries	truly
certain	certainly	changes	clearly	co	must	my	myself	name	namely	try	trying	twice	two	un
com	come	comes	concerning	consequently	nd	near	nearly	necessary	need	under	unfortunately	unless	unlikely	until
consider	considering	contain	containing						new	unto	up	upon	us	use
corresponding	could	couldn't	course						non	used	useful	uses	using	usually
definitely	described	despite	did						not	value	various	very	via	viz
different	do	does	doesn't						obviously	vs	want	wants	was	wasn't
don't	done	down	downwards						ok	way	we	we'd	we'll	we're
each	edu	eg	eight	either	okay	old	on	once	one	we've	welcome	well	went	were
else	elsewhere	enough	entirely	especially	ones	only	onto	or	other	weren't	what	what's	whatever	when
et	etc	even	ever	every	others	otherwise	ought	our	ours	whence	whenever	where	where's	whereafter
everybody	everyone	everything	everywhere	ex	ourselves	out	outside	over	overall	whereas	whereby	wherein	whereupon	wherever
exactly	example	except	far	few	own	particular	particularly	per	perhaps	whether	which	while	whither	who
fifth	first	five	followed	following	placed	please	plus	possible	presumably	who's	whoever	whole	whom	whose
follows	for	former	formerly	forth	probably	provides	que	quite	qv	why	will	willing	wish	with
four	from	further	furthermore	get	rather	rd	re	really	reasonably	within	without	won't	wonder	would
gets	getting	given	gives	go	regarding	regardless	regards	relatively	respectively	wouldn't	yes	yet	you	you'd
goes	going	gone	got	gotten	right	said	same	saw	say	you'll	you're	you've	your	yours
greetings	had	hadn't	happens	hardly	saying	says	second	secondly	see	yourself	yourselves	zero		
has	hasn't	have	haven't	having	seeing	seem	seemed	seeming	seems					
he	he's	hello	help	hence	seen	self	selves	sensible	sent					

1st, 2nd, 3rd Trimester ?

1st, 2nd 3rd degree burn ?

ALONGSIDE	0	PROPERLY	0
CHRONICALLY	9	SIDE	238
CONSISTS	0	SIDED	73
COVERED	0	SPECIFIC	1356
FINDS	0	THINGS	26

10 tokens in IHTSDO list, but not in mySQL list…

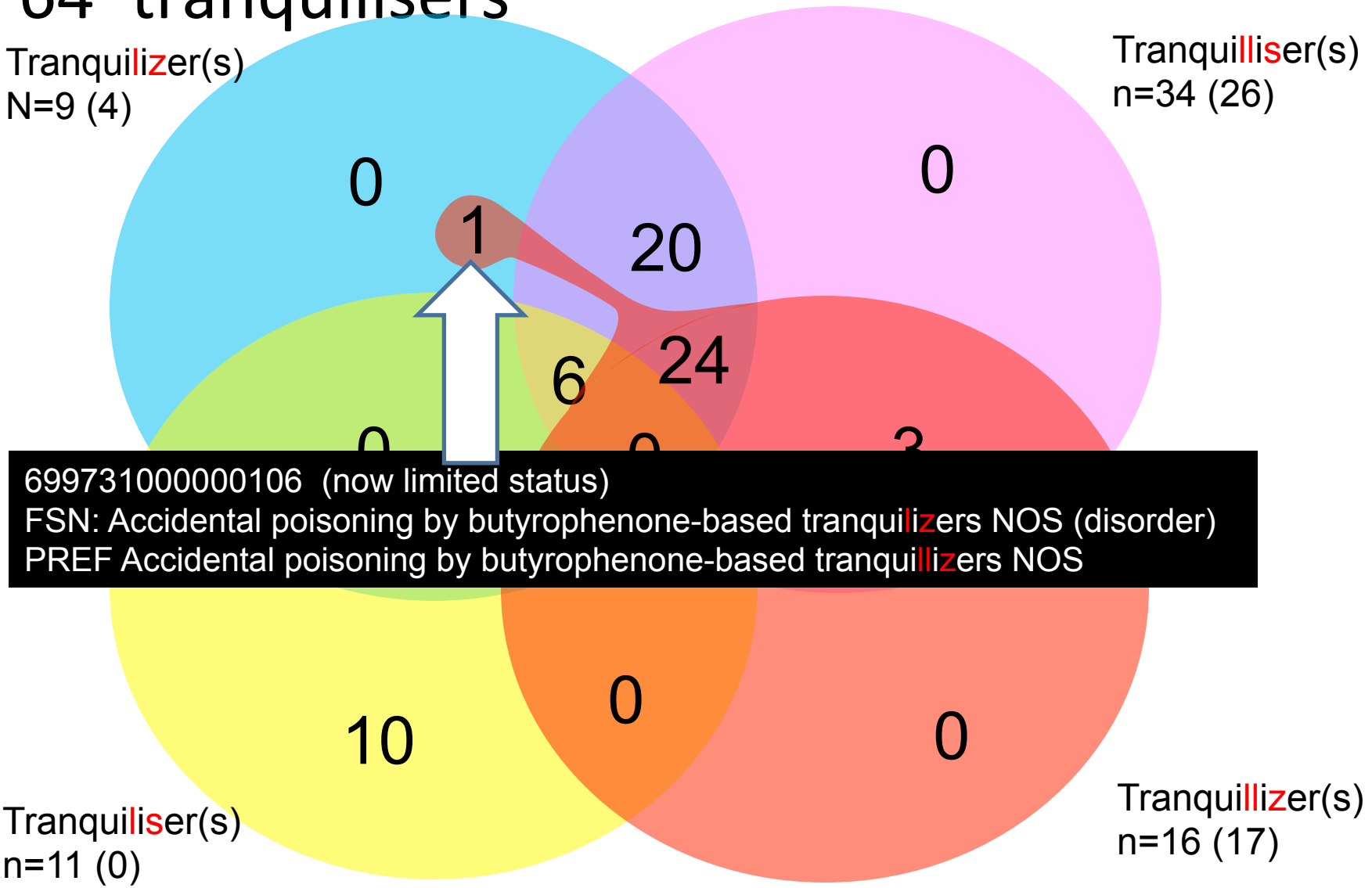
Word Equivalence:
64 ‘tranquilisers’

206129006
FSN: Fetus or neonate affected by maternal tranquilizers during labor and delivery (disorder)
PREF: Fetus or neonate affected by maternal tranquillizers during labor and delivery *enUS*
PREF: Foetus or neonate affected by maternal tranquillisers during labour and delivery *enGB*
SYN: Fetus or neonate affected by maternal tranquilizers during labor and delivery
SYN: Fetus or neonate affected by maternal tranquillisers during labour and delivery

111766000
FSN: Poisoning by tranquilizer (disorder)
PREF: Poisoning by tranquilizer *enUS*
PREF: Poisoning by tranquilliser *enGB*
??? Tranquillizer ???

292343009
FSN: Benzodiazepine sedative adverse reaction (disorder)
PREF: Benzodiazepine sedative adverse reaction
SYN: Adverse reaction to benzodiazepine-based tranquilliser
SYN: Adverse reaction to benzodiazepine-based tranquillizer
??? Tranquilizer ???

Word Equivalence:
64 ‘tranquilisers’



IHTSDO Lexical Resources
Word and Phrase Equivalence

zres_WordEquivalents_en-US_INT_20020731.txt (n=4666)
res_WordEquivalents_en-GB_GB1000000_20060401.txt (n=71)

cicatrix	rupture	perineal	laryngeal	gastric	scrotal
scar	tear	perineum	larynx	stomach	scrotum
erythrocyte	bleeding	scapula	bronchial	pancreas	penile
red blood cell	haemorrhage	shoulder blade	bronchus	pancreatic	penis
gangrene	bruise	leucocyte	mediastinal	ileal	prostate
gangrenous	contusion	white blood cell	mediastinum	ileum	prostatic
avulsed	telangiectasia	platelet	mouth	duodenal	ovarian
avulsion	telangiectasis	thrombocyte	oral cavity	duodenum	ovaries
neurapraxia	cancer	thymic	palatal	rectal	ovary
neuropraxia	malignant neoplasm	thymus	palate	rectum	uterine
injury	malignant tumour	cardiac	roof of mouth	caecal	uterus
trauma	spinal	heart	teeth	caecum	adenohypophysis
agenesis	spine	spleen	tooth	anal	anterior pituitary
aplasia	vertebra	splenic	lingual	anus	neurohypophysis
neoplasm	arm	aorta	tongue	kidney	posterior pituitary
tumour	upper limb	aortic	pharyngeal	renal	adrenal
verruca	carpal	arterial	pharynx	hepatic	suprarenal
wart	carpus	artery	peritoneal	liver	ear drum
boil	wrist	vein	peritoneum	gall bladder	eardrum
furuncle	leg	venous	oesophageal	gallbladder	tympanic membrane
inflammation	lower limb	cranial	oesophagus	testes	bone
inflammatory	thoracic	cranium	bowel	testicle	bony
lump	thorax	skull	intestinal	testicular	osseous
mass	abdomen	nasal	intestine	testis	cartilage
calculus	abdominal	nose	lung		cartilaginous
stone	pelvic		pulmonary		
	pelvis				

IHTSDO Lexical Resources Character Substitutions (1)

DeveloperToolkit\Indexes\tls1_Index_Generator_INT_20100131.zip

\conf\greek_character_conversion_table.txt (2004)

\conf\char_conversion_table.txt (2004)

ALPHA	char=A	0391	char=α	03B1	NU	char=N	039D	char=ν	03BD
BETA	char=B	0392	char=β	03B2	XI	char=Ξ	039E	char=ξ	03BE
GAMMA	char=Γ	0393	char=γ	03B3	OMICRON	char=O	039F	char=ο	03BF
DELTA	char=Δ	0394	char=δ	03B4	PI	char=Π	03A0	char=π	03C0
EPSILON	char=E	0395	char=ε	03B5	RHO	char=P	03A1	char=ρ	03C1
ZETA	char=Z	0396	char=ζ	03B6	SIGMA	char=Σ	03A3	char=σ	03C3
ETA	char=H	0397	char=η	03B7	TAU	char=T	03A4	char=τ	03C4
THETA	char=Θ	0398	char=θ	03B8	UPSILON	char=Y	03A5	char=υ	03C5
IOTA	char=I	0399	char=ι	03B9	PHI	char=Φ	03A6	char=φ	03C6
KAPPA	char=K	039A	char=κ	03BA	CHI	char=X	03A7	char=χ	03C7
LAMBDA	char=Λ	039B	char=λ	03BB	PSI	char=Ψ	03A8	char=ψ	03C8
MU	char=M	039C	char=μ	03BC	OMEGA	char=Ω	03A9	char=ω	03C9

IHTSDO Lexical Resources Character Substitutions (2)

- IHTSDO Diacritic Character Substitutions:

A á â ã ä å ā ă ą ǎ ả À Ä	A Á Â Ã CC04; Å à Ā Ă Ą Ǎ ả
AE æ ǽ	AE Æ
C ç ć ĉ ċ č	C Ç Ć Ĉ Ċ Č
D ď đ ð	D Ď Đ Đ
E è é ê ë ē ĕ ė ę ě	E È É Ê Ë Ē Ĕ Ė Ę Ě
F *f*	
G ĝ ğ ġ ģ	G Ĝ Ğ Ġ Ģ
H ĥ ħ	H Ĥ Ħ
I í ì î ï ĩ ī ĭ į ı	I Í Ì Î Ï Ĩ Ī Ĭ Į I
IJ ij	IJ IJ
J ĵ	J Ĵ
K ķ ĸ	K Ķ
L Í ļ ľ l· ł	L Ĺ Ļ Ľ Ŀ Ł
N ñ ń ņ ň ŉ ŋ	N Ñ Ń Ņ Ň ŉ Ŋ
O ò ó ô õ ö ō ŏ ő ơ ǒ ø	O Ò Ó Ô Õ Ö Ō Ŏ Ő Ơ Ǒ
OE œ	OE Œ
R ŕ ŗ ř	R Ŕ Ŗ Ř
S ś ŝ ş š	S Ś Ŝ Ş Š
T ţ ť ŧ	T Ţ Ť Ŧ
U ù ú û ü ũ ū ŭ ů ű ų ư ǔ ū ǘ ǔ ù	U Ù Ú Û Ü Ũ Ū Ŭ Ů Ű Ų Ư Ǔ Ū Ǘ Ǔ Ù
W ŵ	W Ŵ
Y ý ÿ ŷ	Y Ÿ Ŷ
Z ź ż ž	Z Ź Ż Ž

Diacritic unicode characters in SCT content with no declared unicode decomposition:
øØ ₂ (μ α)

Other issues

Case sensitivity

‘AIDS procedure’ ⇨

Minimum token length

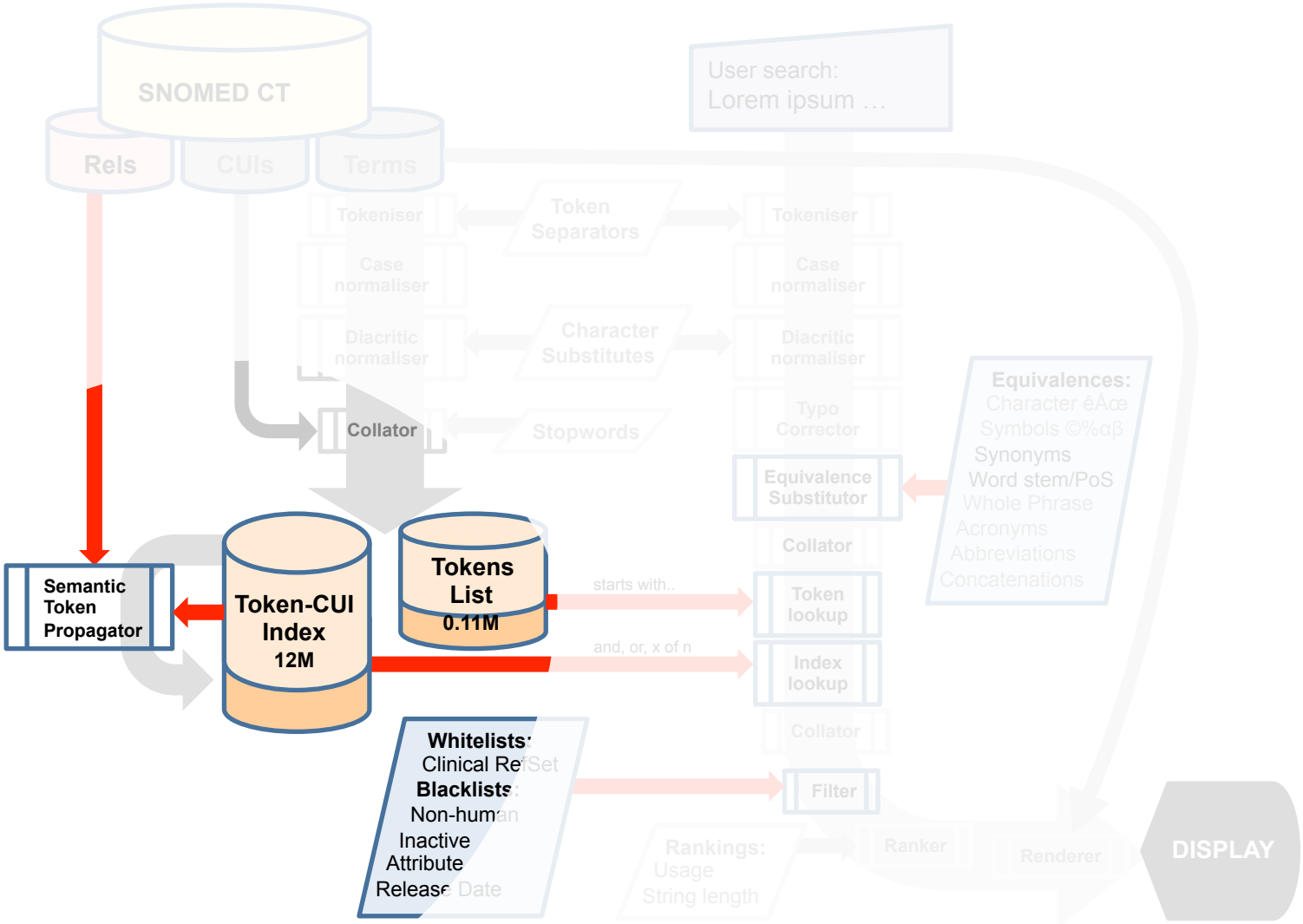
3 char: HIV, WCC etc

2 char: MI, Hb, Fe etc

1 char: Haemoglobin A vs Haemoglobin S

AIDS - Acquired immunodeficiency syndrome antibody test
AIDS (HTLV-III) screening
Aids for severely handicapped
Aids for severely handicapped (procedure)
External memory aids training
External memory aids training (procedure)
External memory aids training (regime/therapy)
Handicapped aids
Handicapped aids (procedure)
Instruction in the use of aids
Notification of AIDS
Notification of AIDS (procedure)
Provision of aids to daily living
Provision of aids to daily living (procedure)
Use of aids for daily life
Use of aids for daily life (procedure)
Use of aids for daily life (regime/therapy)
Use of bathroom aids
Use of bathroom aids (procedure)
Use of bathroom aids (regime/therapy)
Use of indoor mobility aids
Use of indoor mobility aids (procedure)
Use of indoor mobility aids (regime/therapy)
Use of kitchen aids
Use of kitchen aids (procedure)
Use of kitchen aids (regime/therapy)
Use of work aids
Use of work aids (procedure)
Use of work aids (regime/therapy)

Semantic Token Propagation



Outline

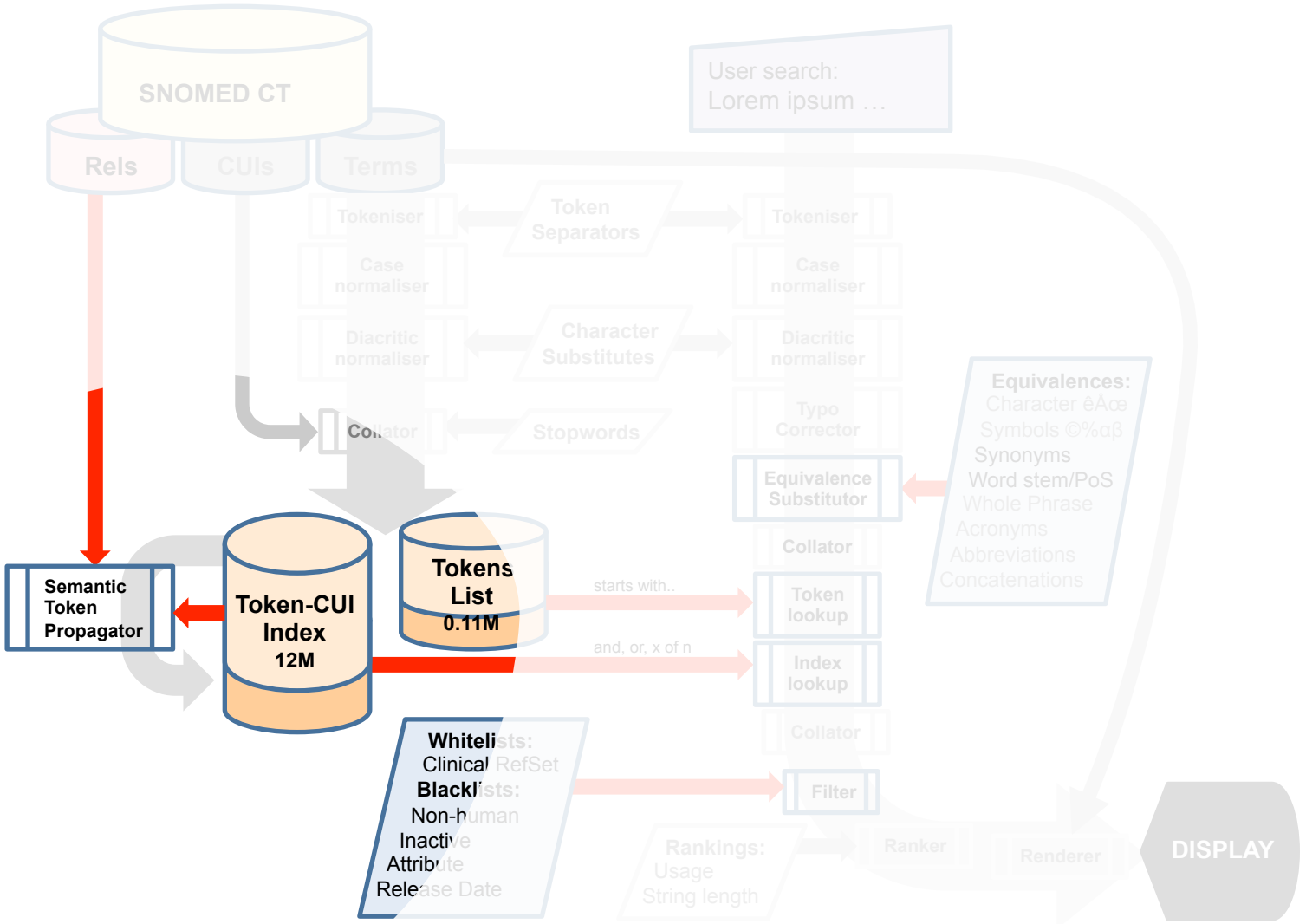
Symptoms

Pathology

Treatment

Evidence

Demonstration : Data Migration Workbench

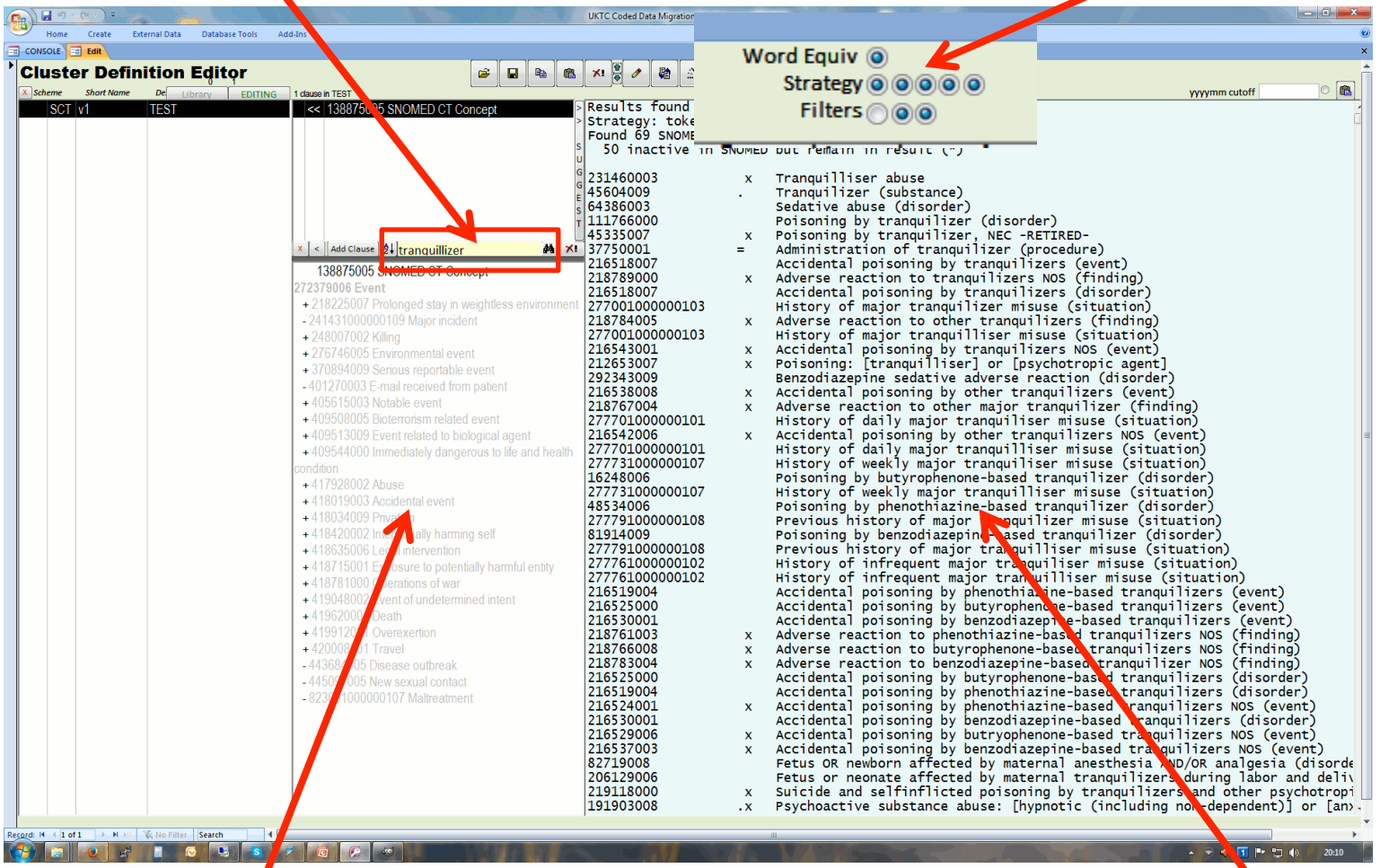


http://www.uktcregistration.nss.cfh.nhs.uk/

Demonstration : Data Migration Workbench

Search expression

Search options



Demonstration : Data Migration Workbench

Word Equivalence

traveller diarrhoea

tranquiliser abuse

flu jab

breast mass

Semantic propagation

acute angina

ossicle removal

defect iris

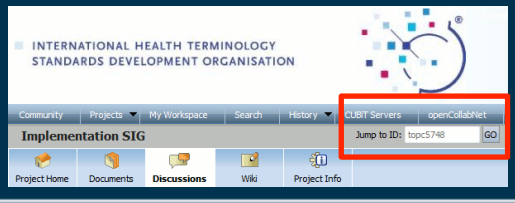
remove ovary fallopian

fat stool

hscic Health & Social Care Information Centre

ThankYou!

jeremy.rogers@hscic.gov.uk

See ImpSIG thread on Collabnet :

Jump to ID **topc5748** (page 2)