

SNOMED CT: Browsing the browsers

Dr Jeremy Rogers

<mailto:jeremy.rogers@nhs.net>

Principal Terminology Specialist, UK Terminology Centre (NHS Connecting for Health)

Olivier Bodenreider

*Staff Scientist, U.S. National Library of Medicine,
National Institutes of Health, Department of Health and Human Services*

With thanks to:

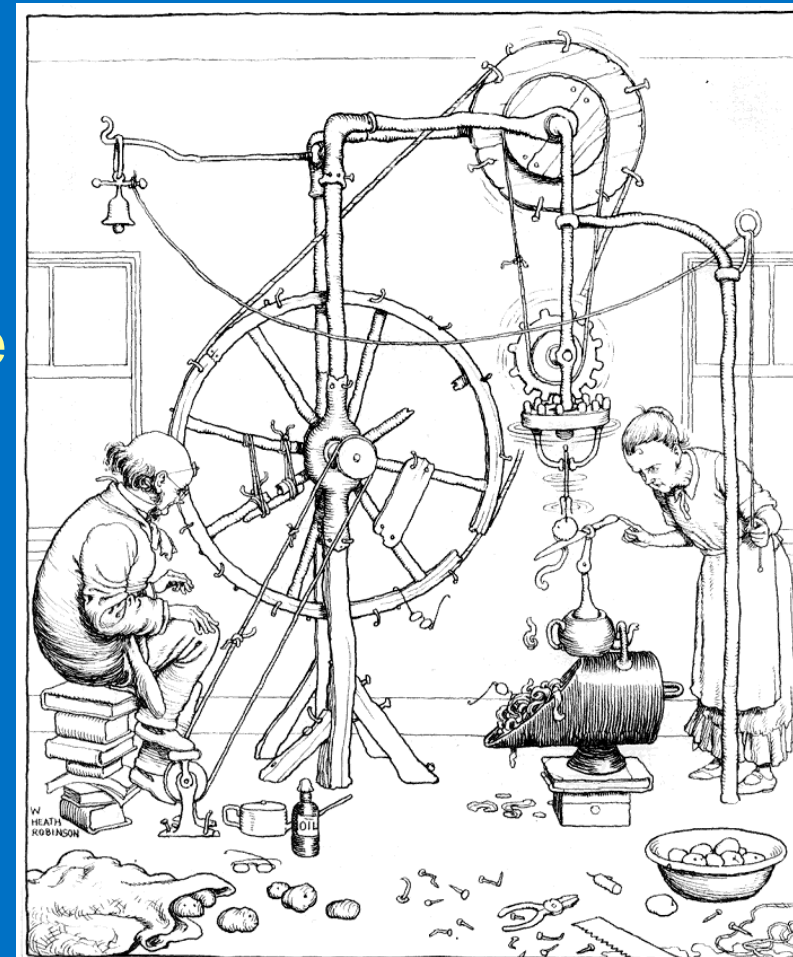
Malcolm Duncan (FDB)

Zac Whitewood-Moores, Lindsey Butler, Amelia Hayselden (NHS UKTC)



Motivation

- SNOMED is complicated
- Needs browsers that:
 - Can make full complexity accessible and understandable
 - Can suppress distracting or unwanted detail
- Impression:
 - Lots of browsers
 - Very different
 - Sometimes poor functionality



The Professor's invention for peeling potatoes.

Goal

- Phase One (in paper):
 - Compile superset of all browsing features/functions implemented by existing SNOMED browsers
 - Initial investigation of featureset variability
- Phase Two (completed - not in paper):
 - Determine reproducibility of featureset testing
 - Document and compare individual browser feature sets
- Purely functional assessment
 - NOT Utility : which browser for which use case?
 - NOT Usability: Ergonomics, performance, stability

- Phase One (in paper)
 - Iterative inspection of browsers
 - Augment feature superset with any novel features
 - Organise emerging feature superset under themes
 - Define reproducible operational tests for each feature
- Phase Two (completed - not in paper)
 - Quadruple re-inspection of all browsers against feature superset
 - Consensus meeting where inspectors disagree
 - Refine operational test definitions

Materials:

19* out of 25 Browsers



Connecting for Health

(AxSys SNOMED Browser)*

CaTTs

CliniClue

CLIVE

EdBrowser

HealthTerm (CareCom)

LExPlorer (HLi)

MyCroft (Apelon)

NCI Terminology Browser

OntoBrowser

OpenKnoME

Protégé-OWL

SNOB

SNOCT (Medsight) **

SnoFlake

TTOC**

UMLS RRF Browser

Virginia Tech Browser

FDB Browser

NOT INSPECTED

ACE (Informatics Inc)

Classification Workbench

Linköping Browser

NHS CUI Demonstrator

Ocean Informatics

Visual Read

*AxSys browser inspected non-interactively via limited screen captures only

** Not in paper

Phase One Results:

Feature superset

- 149 features in browsing superset
 - includes 6 not implemented in ANY browser but could be a good idea
- 8 Themes:
 - Core table data (sct_concepts, sct_relationships, sct_descriptions)
 - Ancillary data (subsets, xmaps, derived views, 3rd Party data)
 - Searching (text-to-concept search and filter)
 - Visualisation (result set sorting, data suppression, tree widgets)
 - Navigation (hyperlinked browsing, favourites)
 - Postcoordination (expression builder, validator)
 - Usability and Interoperation (drag & drop, resultset import/export)
 - Other (content error reporting)

Phase One Results: **Sample**

SEARCHING

Text-to-concept search techniques and filtering

Search by ConceptID	105531004	Concepts can be searched for by direct entry of their conceptID, as well as
Has (and/or does not have) concept status	Motor	Search can be filtered to return ONLY those concepts with status=6 (limited
Has (and/or does not have) concept supercategory	Motor	Search can be filtered to ONLY return 2 concepts that are in [Special Conco
Has (and/or does not have) ancestor = conceptID	Motor	Search can be filtered to show ONLY 1 concept that is a subtype of '415577
Has (and/or does not have) relationship attX with %Y	endoscopic	AND NOT (procedureDevice=endoscope) returns 22 procedures with 'endos
Is (or is not) used in relationship attX from %Y	valve NOT PartOf anat.	(Jan 2008 data) should return 280646001 Valve of nasolacrimal duct, 11055
By namespace or originating authoring centre	Gelfoam	Search can be filtered to return only the 2 concepts that are in the 1000000
By description status	Phytohemagglutinin	Search restricted to retired descriptions (status=1) returns 10669005 ONLY
By description type	insane	Search restricted to preferred term returns zero results (Unrestricted search
By description language	tranquilizer in 'en-US'	(Jan 2008 data) returns 36 concepts without filtering to restrict results to en
Is (not) member of subset	lamisil in Subset:109034	(Jan 2008 UK data) returns 28 active concepts if filtered to include only thos
Exact string match (including on multitoken strings)	[feeding education]	Only 4 CORE concepts are returned on exact match; 6 will be returned by a
Partial string match (wildcarded)	tetral*	Browser returns at least 86299006 Tetralogy of Fallot (may also return other
Force exclude/include terms that match token	tetral* -shunt	Browser does NOT return 12363009 Complete repair of tetralogy of Fallot wi
Force exclude/include matches by frequency of token in index		
Phonetic query expansion (Metaphone, SoundEx etc)	epididimiss	Browser either automatically substitutes correct spelling 'epididymis' or pro
Stemming and part of speech lexical substitute query expansion	toys	Search on 'toys' will return 97 results if the plural is being normalised to 'toy
Strip trailing/leading spaces & non-alpha characters on conceptID	3419005	Any concept ID followed by white space or a comma
Word concatenation query expansion	wheel chair	Browser returns 225612007 (under clinical findings) as well as other results
Colloquial term or external term substitution query expansion	fetal	Browser returns [401091000 Cold agglutinins level : foetal cells (procedure)]
Stopword list (query contraction by blacklist)	between / specified	No results (as opposed to the expected hundreds of results) are returned to single word using
Token separators other than WS in descriptions	DOC	Exact word match returns [131459009 Cham-Doc cattle] and [1336006 11-Deoxycorticostero
Force case sensitive/insensitive searching	ACE vs ace	Case sensitive search on ACE will not NOT also return Ace bandage (16556

Handling of diacritics

Handling of diacritics	descript ing des	Lowchen / Ménière	Browser Browser Complex Browser Search b 430 resu Returns Returns Browser Browser Browser displays search results ordered according to some metric for the a The user can set a threshold maximum number of results to be returned, or Where the browser computes a match accuracy value on each match, the u Result = all concepts returned PLUS all their descendants on July 2007 con
Full regular expression syntax for search expression	%M_ni_re% or equiv		
Type ahead autocompletion / suggestion			
Grouping/hesting of results by supercategory / subsumption	prosthesis*		
Returns matching descriptionID rows	ear AND catheter		
Returns non-redundant set of concepts with matching descriptions	ear AND catheter		
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	ear catheter		

[Lowchen] returns [132607008 Löwchen dog] despite no unaccented match; [Ménière] returns 252546006 despite no accented match

Phase Two Results:

Validation of operation tests



Connecting for Health

- Six browsers quadruple inspected
 - Four reviewers (inc 3 SNOMED content authors) in parallel
 - Consensus meeting where any disagreement
 - Operational tests refined to increase reproducibility
- Solo vs Consensus Results:

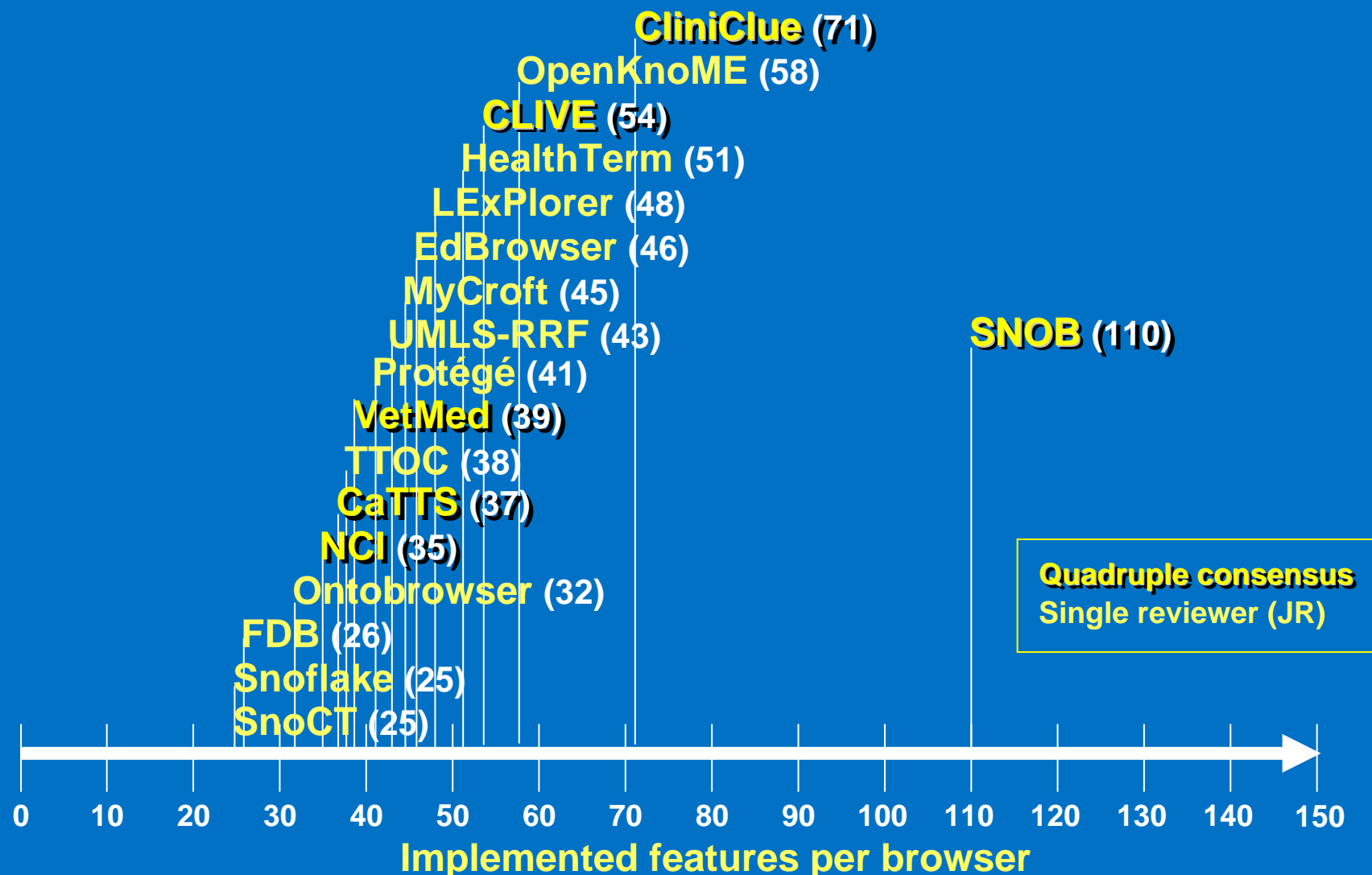
BROWSER	SOLO SCORE	CONSENSUS SCORE
SNOB	110	110
CLINICLUE	68	71
CLIVE	56	54
VETMED	39	39
CATTS	35	37
NCI	36	35

Thick client / standalone

Web based browsers

Phase Two Result:

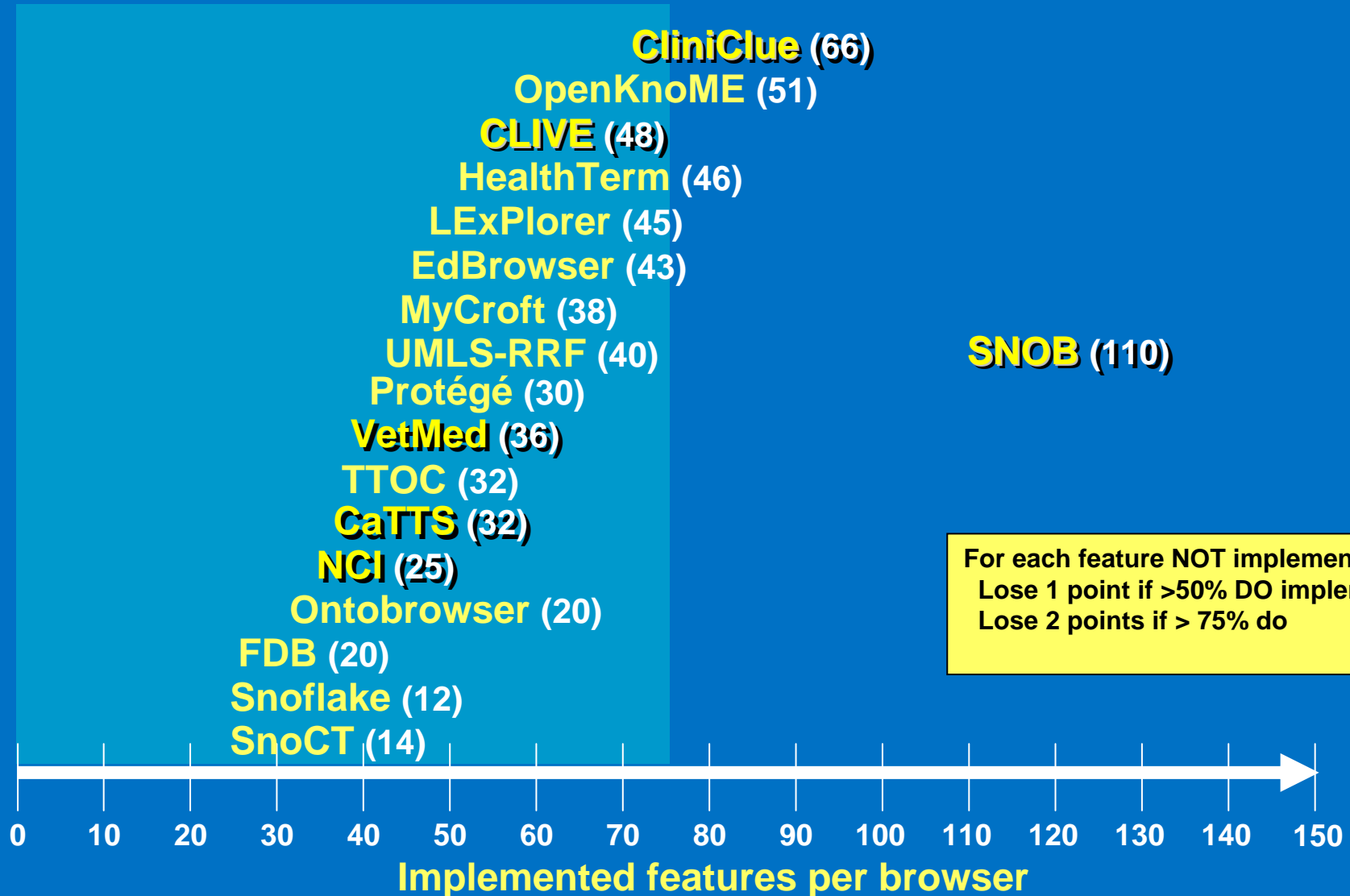
Per-browser feature count



Phase Two Result: ...with negative marking



Connecting for Health



Phase One Results:

Impoverished featuresets

- Average of only 40 features per browser
 - Range 18-110, SDEV=17
 - About a quarter of what's possible
- All but one browser implements *less than half* all possible features
 - Where (almost) all of what's possible is implemented at least one other browser

Phase One Results:

Idiosyncratic featuresets

- Only 5 features found in all browsers
- Only 17 more found in most (>66%)
- 70 features in more than one browser, but only a minority of all browsers (<33%)
- 19 further features unique to one browser
 - But not the same browser
 - 6 browsers have at least one unique feature

How idiosyncratic, exactly ?



Featureset comparison

Core table data

CORE SNOMED CT DATA

Basic Concept Table Info

Fully Specified Name	Y	Y	Y	Y	Y	Y	Y		Y		Y	Y	Y		Y	Y	Y	Y
ConceptID	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ConceptStatus (original numerical value)	Y					Y	Y		Y			Y						
ConceptStatus (human readable text - current, ambiguous etc)	Y	Y		Y	Y			Y						Y		Y	Y	
IsPrimitive	Y	Y		Y	Y	Y	Y		Y	Y	Y	Y				Y		
CTV3ID	Y	Y		Y	Y	Y	Y	Y	Y			Y	Y		Y		Y	Y
SNOMEDRT ID	Y	Y		Y	Y	Y	Y	Y	Y			Y	Y		Y			Y

Basic Description Table Info

DescriptionID	Y	Y		Y		Y	Y		Y		Y	Y						
ConceptID	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y		Y	Y
DescriptionStatus (original numeric value)	Y					Y			Y			Y						
DescriptionStatus (current, erroneous, retired etc)	Y			Y														
DescriptionType (preferred, synonym, FSN)	Y	Y		Y	Y	Y	Y		Y		Y	Y	Y				Y	Y
Term	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
InitialCapitalStatus Flag Value Displayed	Y			Y					Y			Y						
Language (en, en-gb, en-us etc)	Y			Y		Y			Y	Y		Y						Y

Basic Relationship Table Info

RelationshipID	Y					Y	Y		Y									
ConceptID1 (or FSN/Preferred term)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
RelationshipType (or FSN/Preferred term)	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y
ConceptID2 (or FSN/Preferred term)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Characteristic Type (defining vs optional vs additional)	Y	Y		Y		Y	Y	Y		Y	Y			Y				
Refinability	Y	Y		Y		Y						Y						
RoleGroup	Y	Y		Y	Y		Y			Y	Y		Y					Y

Featureset comparison

Ancillary data



Connecting for Health

ANCILLARY. 3rd PARTY AND DERIVABLE DATA

Ancillary Table Info

Crossmaps	Y	Y		Y		Y			
Subset membership	Y								
Concept history (when added etc)	Y							Y	
Namespace/extension identification	Y			Y					Y

3rd Party Info

Original SNOMED ID Preferred Term	Y								
Original CTV3 Preferred Term	Y								
Original external scheme terms	Y	Y		Y			Y		
Hyperlink to Xmapped scheme browsers	Y						Y		
Allowable Concept Model Relationships	Y		Y						Y
Dynamic flagging of modelled relationships outside concept model	Y								

Derivable Table Info

All concepts that use X in their definition	Y	Y	Y	Y	Y			Y	Y	
Forward history relationships	Y	Y			Y	Y	Y	Y		
Reverse history relationships	Y			Y	Y		Y			Y
Content Metrics	Y							Y		
Reverse cross maps	Y						Y			Y
Code conversion (single SCTID -> nearest target scheme map)									Y	
All members of one subset	Y				Y				Y	
Short normal form	Y							Y		
Long normal form	Y	Y								
Dynamic inheritance of relationships				Y	Y	Y				Y
Similarly indexed concepts										Y
Other concepts with exact same description(s)										Y

SNOMED Release Versioning

Names of all source files actually loaded	Y									
Currency of content (whether release is outdated)	Y	Y						Y		Y
Simultaneous browsing of any set of different SCT releases	Y	Y	Y							Y

Featureset comparison

Visualisation & Navigation



Connecting for Health

VISUALISATION

Hierarchy And Other Content Visualisation

Indefinitely Expandable/Collapsible Ancestor Tree	Y		Y					Y		Y		Y	Y	Y	Y		Y	
Indefinitely Expandable/Collapsible Descendant Tree	Y	Y	Y	Y	Y			Y	Y	Y				Y	Y	Y	Y	Y
Indefinitely Expandable/Collapsible Bidirectional Polyhierarchy	Y		Y											Y				
Bidirectional transitive closure only browser	Y		Y															
Descendant Hierarchy does not have to start from SNOMEDCT Concept	Y	Y	Y	Y		Y	Y	Y					Y	Y		Y		Y
Large result warning and bail-out	Y	Y		Y					Y					Y				
Flag where same concept currently displayed >1 time same			Y															
Alphasorting of synonyms	Y	Y	Y			Y	Y	Y	Y					Y				
Alphasorting of siblings	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y
Collation of relationships on same attribute within a role group	Y	Y	Y	Y	Y	Y		Y					Y				Y	
Relationships are not sorted (+/- collated) numerically by attribute	Y		Y	Y		Y	Y	Y					Y	Y	Y	Y	Y	Y
Relationships are alphasorted (+/- collated) by role name	Y		Y	Y		Y		Y					Y				Y	
Sets of same relationships can't be sorted numerically by	Y		Y	Y	Y	Y	Y	Y	Y				Y	Y		Y	Y	Y
Alphasorting within sets of same relationship by value name	Y		Y	Y				Y						Y			Y	
Conceptual grouping or filtering of relationships by relationship supertype																		
Rendering of superscript/subscript descriptors																		
Rendering of diacritics	Y	Y						Y	Y	Y			Y			Y	Y	
Colour and typographic coding of hierarchy nodes by status etc	Y	Y	Y										Y					
Snoflake / Cloud / Other GUI views			Y				Y	Y				Y						Y
Hierarchies on non IS-A relationships				Y														

Dynamic Display Filtering

By concept status			Y							Y								
By description status	Y							Y						Y				
By description type			Y							Y								
By description language			Y															

NAVIGATION

Hierarchy Navigation

Browse history (back, forward buttons)	Y					Y	Y		Y	Y	Y	Y	Y		Y			
Refocus browser on any displayed concept (hyperlinked browsing)	Y	Y	Y			Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y
Refocus browser on 'Home' concept	Y	Y				Y	Y				Y	Y	Y	Y				
Persistent user-defined favourites list	Y		Y	Y				Y				Y		Y				Y

Featureset comparison

Usability & Interoperation

USABILITY & INTEROPERATION

User Interface Usability and Interoperability

Internal Drag-and-drop	Y	Y	Y					Y				Y	Y		Y
External Drag-and-drop (e.g. to MS Word, Outlook)	Y	Y		Y											
Scroll-wheel enabled whenever any long list visible (e.g. tree view)	Y		Y	Y	Y				Y	Y	Y	Y	Y	Y	Y
Maximum single session hierarchy browsers	>9	2	>9	>9	>9	>9	1	>9	1	>9	>9	1	1	1	>9
Hierarchy browser expansion state memory			Y		Y			Y	Y	Y					Y
Concept information widget state memory	Y	Y						Y	Y				Y	Y	
Adjustable font size	Y	Y	Y		Y	Y			Y	Y	Y	Y	Y	Y	Y
Browser component slaving on same or different databases	Y		Y						Y						
Subsumption hierarchy only view	Y	Y	Y	Y	Y			Y		Y				Y	
Plug-in user extensible architecture / API	Y	Y								Y					
Single Application Window Interface	Y	Y			Y	Y	Y		Y	Y	Y		Y	Y	Y
MDI Container Interface					Y										Y
Multiple spawnable window interface	Y		Y		Y			Y				Y		Y	Y
Independently scalable windows	Y		Y	Y	Y			Y				Y		Y	Y
User configurable suppression of IDs in tree view browser	Y	Y	Y	Y			Y								Y
Copy hierarchy as text	Y	Y	Y									Y	Y	Y	
Copy hierarchy as graphic	Y														
Copy hierarchy as XML / HTML / RTF		Y	Y									Y		Y	Y

Featureset comparison

Postcoordination, Other



Connecting for Health

POSTCOORDINATION

Postcoordinated Expressions

Expression builder UI	Y	Y	Y	Y			Y
Allowable choices offered predictively	Y		Y				
Refinable choices offered predictively	Y	Y	Y	Y			
Refinability flag aware	Y	Y					
Allowable/Refinable/Optional candidate forecasting		Y					
Nested expression building							Y
Role group aware	Y	Y			Y		
Postcoordinated expression validation against concept model				Y			Y
Dynamic classifier				Y			Y

MISCELLANEOUS

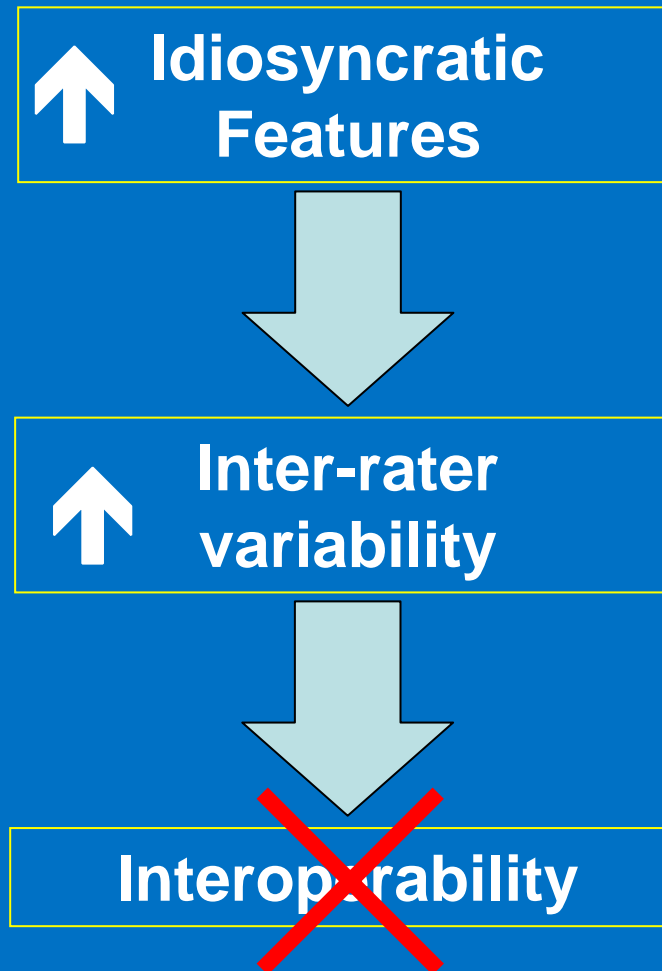
Content error reporting

Internal content error reporting	Y		Y
Internal content error management	Y		Y
External content error reporting	Y		Y
External content error syndication	Y		Y

Other User Interface Features

Subset editor and exporter	Y			Y			
TQL' Parser			Y				Y
Standalone (does not require internet connection)	Y	Y	Y		Y	Y	Y
Bail-out to view on raw tables						Y	
Distributed independent of data (user can rebuild)	Y		Y		Y	Y	Y

Why this matters



Idiosyncratic searching...



Summary

- Growing number of SNOMED CT browsers
 - One best of breed by clear margin
- Generally limited and idiosyncratic sophistication
 - All but 1 implements <50% possible functionality
 - Poor overlap in featuresets
 - Browsing of ancillary data, and postcoordination expression builders, are particularly limited
 - Web-based browsers generally very poor
- Significant risk to interoperability
 - Especially (but not only) search/sort behaviours

Future work

- Maintain feature superset
 - New browsers, new features
 - E.g. multilingual user interface
- Test and Compliance Data
 - Existing operational tests exploit chance features of current international and UK release data
 - These are transient; tests not portable across releases
 - Requirement for stable test and compliance extension
 - Alpha version completed
- Minimum browser requirements



Connecting for Health