

Some (Sketchy) Case Studies

Bijan Parsia

COMP60462

March 17, 2009

Big Terminologies

- Taxonomy vs. Ontology
 - What are taxonomies used for?
 - Controlled vocabularies
 - Faceted navigation
 - Data integration
 - How to develop/maintain a big terminology?
 - Ontologies can help
 - Ontology for development
- For example, SNOMED-CT
 - And GO, Galen, FMA, NCI Thesaurus, etc. etc. etc.

NCI Cancer Thesaurus

- It's big
 - 50,000 concepts
 - Team of up to 20
- Provide terminology for applications
 - [Cancer Image Database](#)
 - Cancer Data Standards Repository
- Heavy investment in tool infrastructure
 - Most recently explanation and modularity services and incremental reasoning
 - Critical to both speeding up development and improving quality



TO LOCATE A TERM: 1 Enter a word or phrase in SEARCH or 2 Navigate to the term in the DATA TREE

SEARCH: Entry

Enter Search Word or Phrase Below

Search:

Match Entire Search Terms Only

SEARCH: Results







Results of Your Search Will Appear Below

Entering a search term above will perform a search for the term entered as well as alternate meanings.

For example, entering 'throat' will search for 'pharynx' as well as 'throat'.

DATA TREE

Browse DataTree Directly, Or Use Search for Prompting

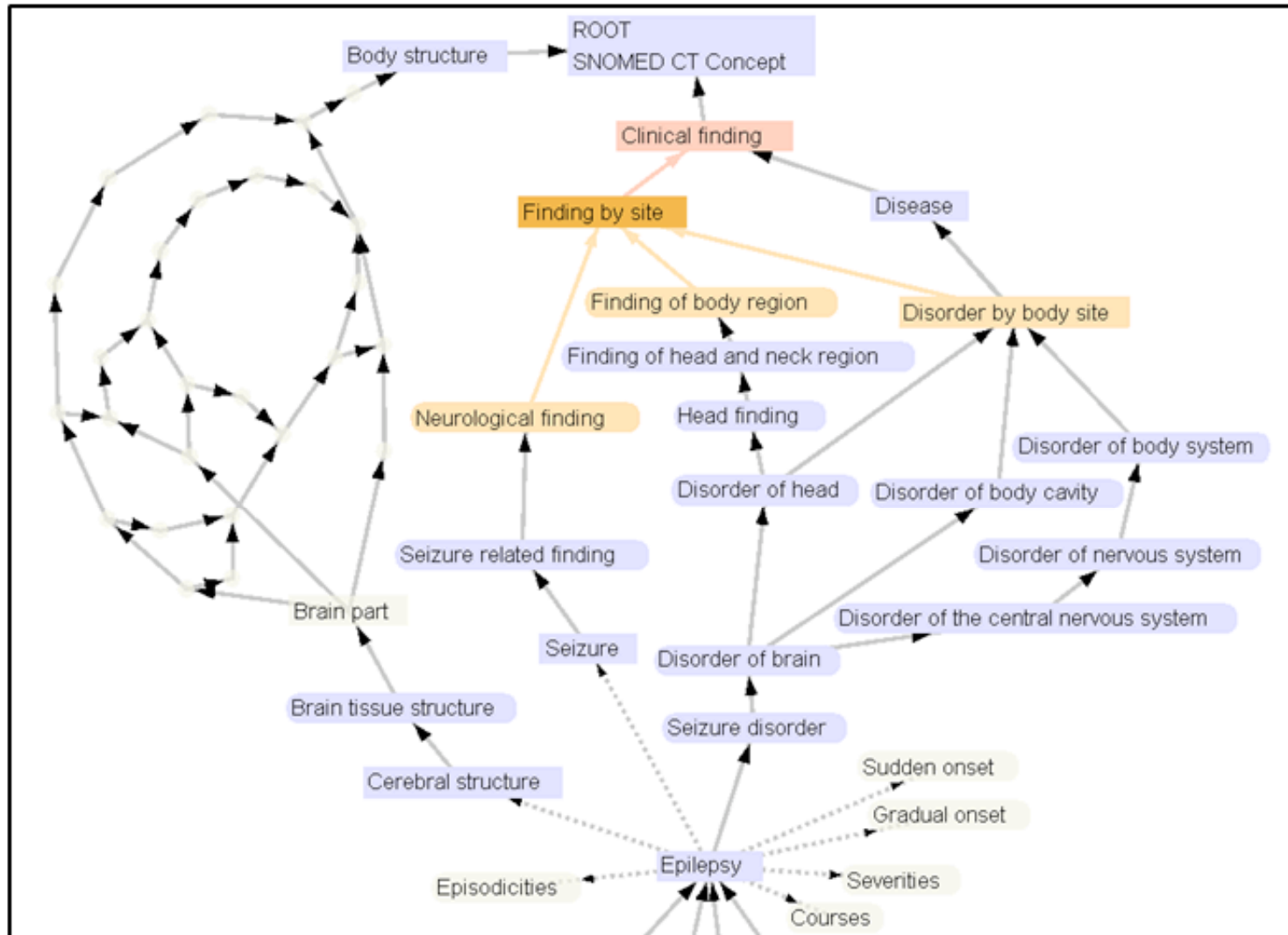
-  [Murine_Tissue_Type](#)
 -  [Body_Fluids_and_Substances_MMHCC](#)
 - [Discharge_MMHCC](#)
 - [Exudate_MMHCC](#)
 - [Mucus_MMHCC](#)
 - [Pleural_Fluid_MMHCC](#)
 - [Sebum_MMHCC](#)
 - [Sputum_MMHCC](#)
 - [Stool_MMHCC](#)
 - [Urine_MMHCC](#)
 -  [Cardiovascular_System_MMHCC](#)
 -  [Digestive_System_MMHCC](#)
 -  [Embryonic_Tissue_MMHCC](#)
 - [Embryonic_Tissue_MEF_MMHCC](#)
 -  [Endocrine_Gland_System_MMHCC](#)

Patient Data Management

- UK NHS £6.2 billion IT program
 - “Connecting for Health”
- Key component: Care Records Service (CRS)
 - “interactive patient record service accessible 24/7”
 - Patient data distributed across local centers in 5 regional clusters, and a national DB
 - Detailed records held by local service providers
 - Diverse applications support radiology, pharmacy, etc
 - Applications exchange messages containing “semantically rich clinical information”
 - Summaries sent to national database
- SNOMED-CT ontology is a key part!
 - provides common vocabulary for data

SNOMED

- Very Large:
 - 373,731 concepts & over 1 million terms
 - NHS version extended to 542,380 classes with
 - 19,828 additional named classes
 - 148,821 class drug taxonomy
- OWL reasoner (FaCT++) classified it
 - Able to classify whole ontology in <4 hours
 - Interesting results come from 19,828 additional named classes
 - 180 missing subClass relationships were found, e.g.:
 - Periocular_dermatitis subClassOf Disease_of_face



Overview

Horizons

- Fetch 0 -5
- Label 0 -5
- Cluster 0 -5

Details

Label/Relation	Value/Ta
ConceptStatus	Current
FullySpecifiedNa...	No family his
ID_KEY	160274005
IsPrimitive	Primitive
Finding context (at...	Known abse
Subject relationsh...	Person in the
Is a (attribute)*	No relevant f

0. SNOM. concepts +dia* New + - 500

Post Coordination

- Vocabulary is extensible at point of use: “post coordination”
 - Users (e.g. clinicians) may enter novel class descriptions
 - Terminology service (reasoner) used to classify descriptions
- Typical new term might be “Allergy caused by almond”
 - Classified as a kind of “nut allergy”
 - Extensible, but coherent

Galen

- Ontology underlying a clinical system
 - Not just a control vocabulary
 - Clinical *forms*

The screenshot shows the Galen clinical system interface for 'Angina pectoris'. The interface is organized into several sections:

- Topics:** 'Angina pectoris' is selected.
- Descriptors:**
 - Presence:** Buttons for 'present', 'possible', and 'excluded'. 'present' is selected.
 - Duration:** A spinner box with '1' and a dropdown menu with 'days'.
 - Severity:** Buttons for 'mild', 'moderate', and 'severe'. 'moderate' is selected.
 - Progress:** Buttons for 'better', 'same', and 'worse'. 'worse' is selected.
- Associated Symptoms:** A button labeled 'all...'.
 - Chest pain:** Buttons for 'absent', 'present', and 'more...'. 'present' is selected.
 - Arm pain:** Buttons for 'absent', 'present', and 'more...'. 'absent' is selected.
- Further History:** Buttons for 'FH', 'PMH', and 'SH'.
- Examinations:** A button labeled 'all...'.
 - Examination:** A button labeled 'normal'.

Form explosion

- The PROMIS system was discontinued in 1987 at 70K forms and accelerating
- A supplier recently expanded all XML forms for data capture in one clinic to 10^7
- The exploding bicycle
(codes for injuries involving cyclists)

Types of Bicycle injury

- 1972 ICD-9 (E826) 8
- READ-2 (T30..) 81
- READ-3 87
- 1999 ICD-10

Taming the explosion

- Break forms into parts
 - Build separate hierarchies
 - *With definitions*
- The reasoner assembles the parts
 - On demand
 - While validating the definitions