Biomedical Semantics in the Big Data Era

Workshop of IMIA WG 6

'Language and Meaning in Biomedicine' (LaMB)

MEDINFO 2015 – São Paulo,Brazil

Speakers

- Tomasz Adamusiak (Thomson Reuters, Boston, MA, USA)
- Ronald Cornet (Academic Medical Center University of Amsterdam, Amsterdam, The Netherlands & Linköping University, Linköping, Sweden)
- Jianying Hu (IBM T. J. Watson Research Center, Yorktown Heights, NY, USA)
- Stephane Meystre (University of Utah, Salt Lake City, Utah, USA)
- Patrick Ruch (University of Applied Sciences Western Switzerland, Geneva, Switzerland)
- Stefan Schulz (Medical University of Graz, Graz, Austria)

Comment Stefan: I removed academic degr because of difficult comparability across countries

Agenda

- Introduction – Ronald Cornet

1. From free text to ontology – Stephane Meystre
2. Bridging natural and formal languages for representing knowledge and information – Stefan Schulz
3. Deep question-answering for biomedical decision support – Patrick Ruch
4. Feature extraction for predictive modeling – Jianying Hu
5. Semantic technology for knowledge discovery – Tomasz Adamusiak

- Overall discussion – all
- Round-up

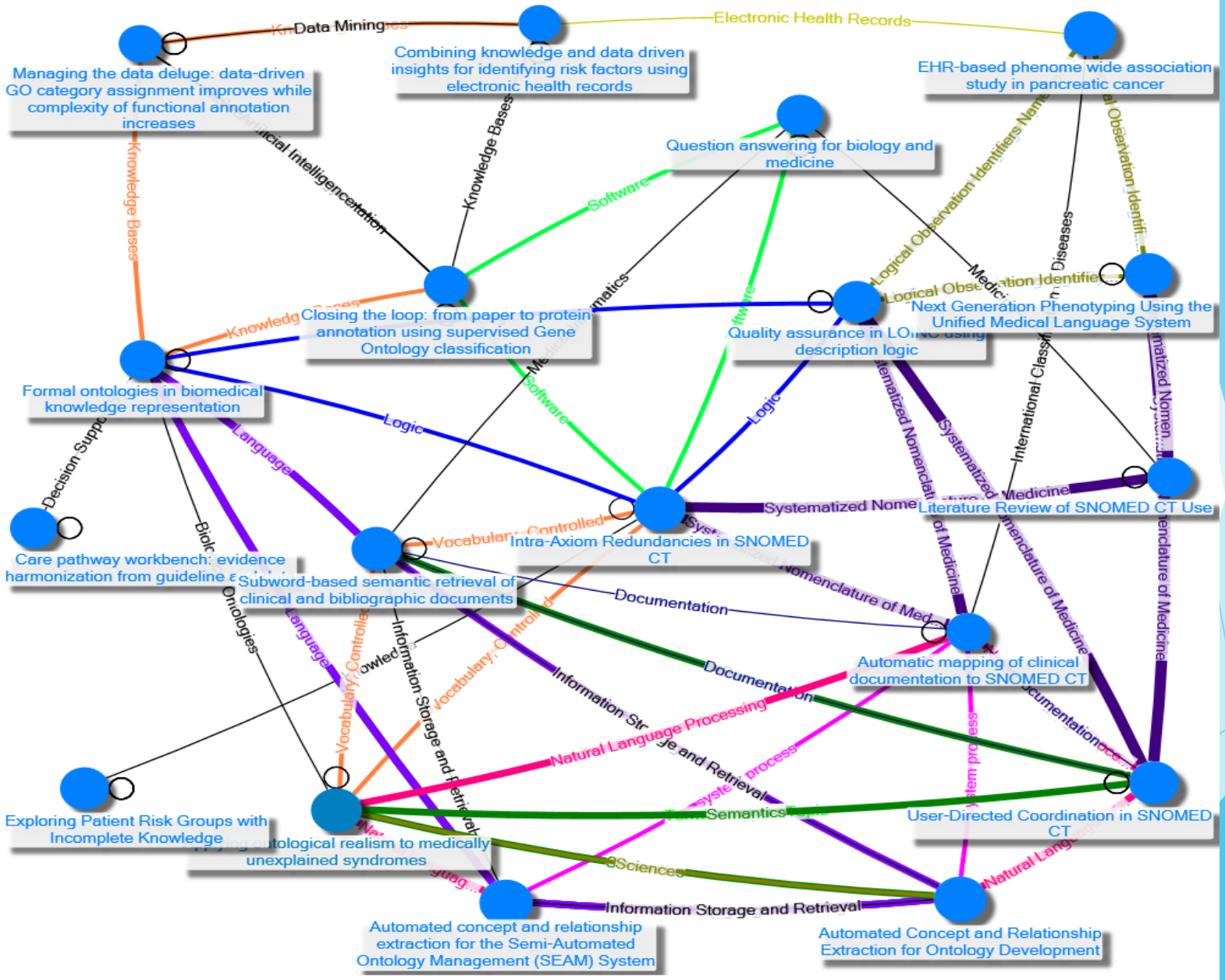
Topics to be discussed

- Applying ontological realism to medically unexplained syndromes
- Automated Concept and Relationship Extraction for Ontology Development
- Automated concept and relationship extraction for the Semi-Automated Ontology Management (SEAM) System
- Automatic mapping of clinical documentation to SNOMED CT
- Care pathway workbench: evidence harmonization from guideline and data
- Closing the loop: from paper to protein annotation using supervised Gene Ontology classification
- Combining knowledge and data driven insights for identifying risk factors using electronic health records
- EHR-based phenome wide association study in pancreatic cancer
- Exploring Patient Risk Groups with Incomplete Knowledge
- Formal ontologies in biomedical knowledge representation
- Intra-Axiom Redundancies in SNOMED CT
- Literature Review of SNOMED CT Use
- Managing the data deluge: data-driven GO category assignment improves while complexity of functional annotation increases
- Next Generation Phenotyping Using the Unified Medical Language System
- Quality assurance in LOINC using description logic
- Question answering for biology and medicine
- Subword-based semantic retrieval of clinical and bibliographic documents
- User-Directed Coordination in SNOMED CT

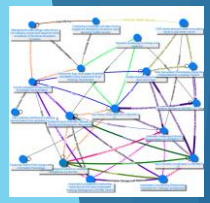
Key Concepts

- Algorithms
- Artificial Intelligence
- Automatic Data Processing
- Bayes
- Bibliometrics
- Biological Ontologies
- Biological Science Disciplines
- Biology
- Case Management
- Classification
- Computational Biology
- Computer Systems
- Computers
- Critical Pathways
- Current Procedural Terminology
- data
- Data Collection
- Data Mining
- Databases, Bibliographic
- Databases, Genetic
- Decision Support Systems, Clinical
- Decision Support Techniques
- Documentation
- Electronic Health Records
- Gene Ontology
- Healthcare Common Procedure Coding System
- Information Science
- Information Storage and Retrieval
- Information Systems
- Intelligence
- International Classification of Diseases
- knowledge
- Knowledge Bases
- Language
- Linguistics
- Logic
- Logical Observation Identifiers Names and Codes
- Meaningful Use
- Medical Informatics
- Medical Informatics Applications
- medical record
- Medical Record Linkage
- Medical Records Systems, Computerized
- Medicine
- MEDLINE
- Molecular Sequence Annotation
- Multilingualism
- Natural Language Processing
- predictive model
- Programming Languages
- Publishing
- PubMed
- Quality Control
- Quality Improvement
- RISk Group Analysis
- ROC Curve
- RxNorm
- Science
- Search Engine
- Semantics
- Software
- system process
- Systematized Nomenclature of Medicine
- Systems Biology
- Terminology as Topic
- Unified Medical Language System
- Vocabulary, Controlled
- Workflow

Topics and key concepts



First presentation



Biomedical Semantics in the Big Data Era

Stefan Schulz

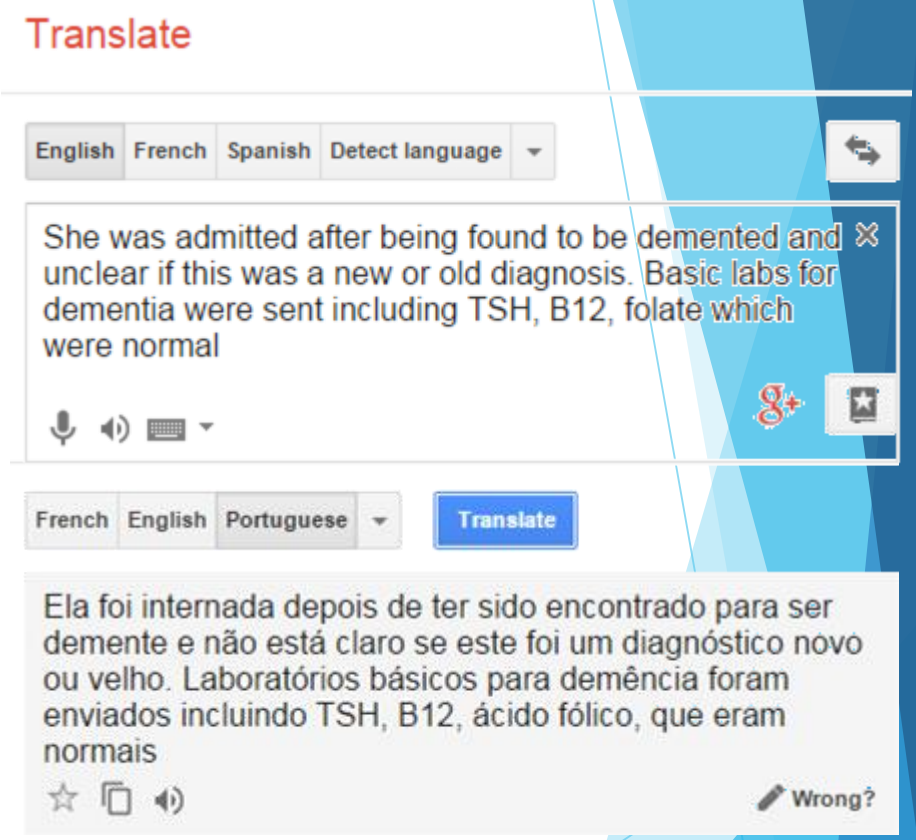
Bridging natural and formal languages for representing knowledge and information

Workshop of IMIA WG 6
'Language and Meaning in Biomedicine' (LaMB)
MEDINFO 2015 – São Paulo,Brazil

The promise of Big Data (?)

- "*Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all. (...) Forget taxonomy, ontology, and psychology.*"

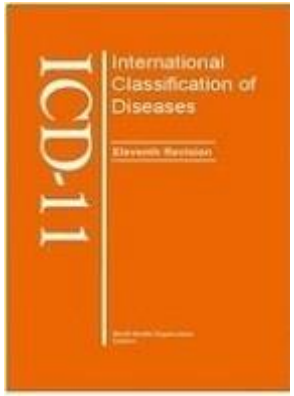
Chris Anderson *Wired Magazine, 2008*

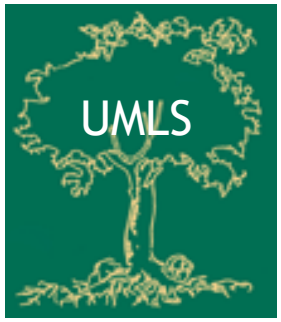


The rise of semantic standards and specifications

Ontologies Terminologies Information Models

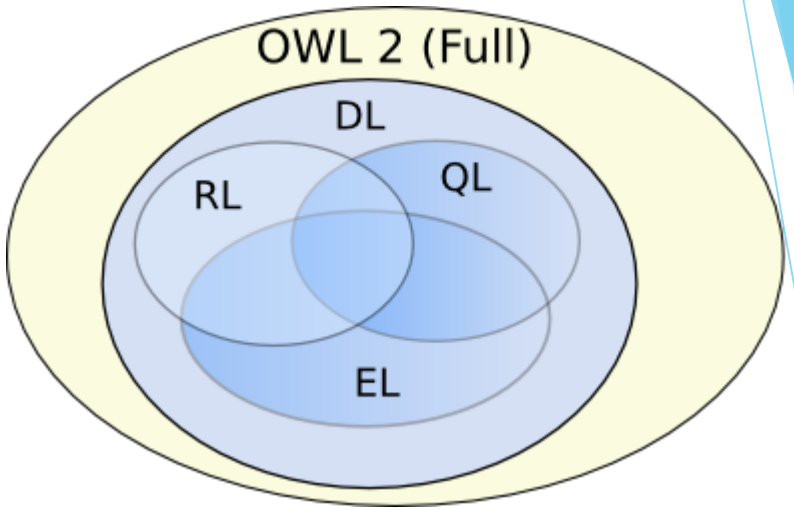
IHTSDO delivering
SNOMED CT®
the global clinical terminology





openEHR

Standardised representation and reasoning formalisms

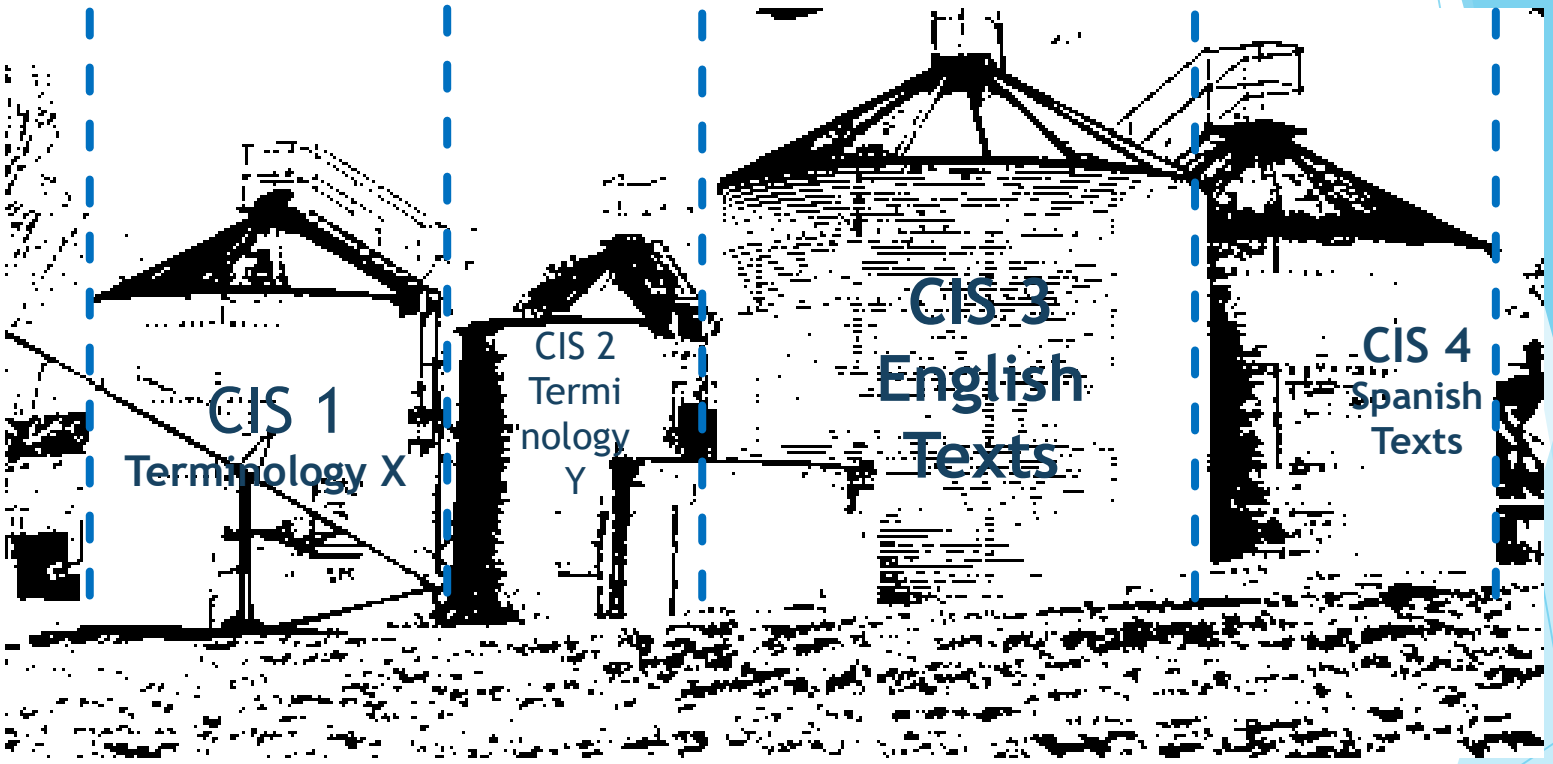


ObservationResult and **isAboutQuality** only (*MassIntake*
and **inheresIn** some *CigaretteTobaccoSmokingSituation*
and **projectsOnto** some (*ValueRegion* and
isRepresentedBy only
(**hasInformationAttribute** some *perDay* and
hasValue some int[>=10])))
SubClassOf *InformationItem* and **isAboutSituation**
HeavyCigaretteTobacco Situation

Persistence of free text narratives in EHRs

Patient is a 80 y/o female with hx of CAD, DM, HTN, left PICA stroke who presented to the ED after a fall. SHe was admitted after being found to be demented and unclear if this was a new or old diagnosis. Basic labs for dementia were sent including TSH, B12, folate which were normal. MRI revealed a mennigioma and old PICA infarct. Likely diagnosis is Alzheimer's Disease. She was started on donepezil and Quetiapine. PT/OT evaluated her and felt that she was safe to be d/c home with services.

Persistence of data silos: No interoperability

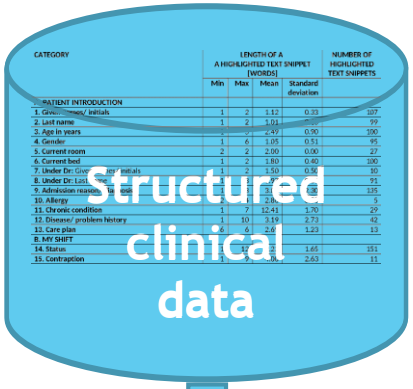


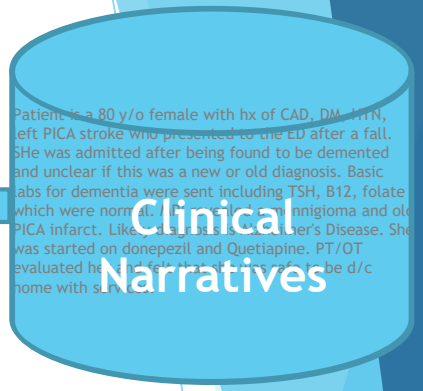
- Proprietary vocabularies / data dictionaries
- Proprietary information templates
- Different natural languages
- Legacy systems that obviate data exchange

Barriers to Semantic Interoperability

- Vocabularies, ontologies, information models:
  - Conflicting and overlapping models of meaning and use
  - Lack of ontological grounding
  - Confuse and ambiguous naming
- Representation and reasoning mechanisms
  - Difficult to learn and to apply
  - Performance issues, computational complexity
  - Lack of industry-standard tools
- Natural language content
  - Idiosyncratic language (abbreviated, ungrammatical)
  - Context dependence

The vision of Semantic Big Data





NLP Pipeline

mining

Information model #1

Information model #3

Information model #3

Virtual Homogeneous Data

Clinical queries
Business Intelligence
Decision Support
Content Summarization
(…)

grounding

Simple upper-level model of disjoint classes

Information Entity

Clinical Entity

- record artefact
- time
- certainty
- provenance
- quality

binding

- life phase (finding)
- procedure
- substance
- observable
- device
- role
- organism

Clinical Termin-ologies Ontologies

Overall discussion

Round-up