Tutorial T5

The Unified Medical Language System (UMLS) and the Semantic Web

Olivier Bodenreider

Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Outline

- Information integration in biomedicine
  - Some issues: naming, normalization, mapping
  - Semantic Web perspective
- Terminology integration in biomedicine
  - Unified Medical Language System
- Some differences between UMLS and SW
Information integration in biomedicine

Some issues: naming, normalization, mapping
Many biomedical entities have several names (synonymy)

- Drug names
- Gene names
- Disease names
- ...

A given name may refer to several different entities (polysemy)

- Nail (body part)
- Nail (medical device)
## Brand names for paracetamol (acetaminophen)

![List of paracetamol brand names](http://en.wikipedia.org/wiki/List_of_paracetamol_brand_names)

<table>
<thead>
<tr>
<th>Brand name</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acamol</td>
<td>Israel</td>
</tr>
<tr>
<td>Atamel</td>
<td>Venezuela</td>
</tr>
<tr>
<td>Adol</td>
<td>Oman</td>
</tr>
<tr>
<td>Aldolor</td>
<td>Israel</td>
</tr>
<tr>
<td>Alvedon</td>
<td>Sweden</td>
</tr>
<tr>
<td>APAP</td>
<td>Poland</td>
</tr>
<tr>
<td>Benuron</td>
<td>Portugal, Germany</td>
</tr>
<tr>
<td>Biogesic</td>
<td>Philippines</td>
</tr>
<tr>
<td>Buscapina</td>
<td>Argentina</td>
</tr>
<tr>
<td>Cemol</td>
<td>Thailand</td>
</tr>
<tr>
<td>Crocin</td>
<td>India</td>
</tr>
<tr>
<td>Dafalgan</td>
<td>Belgium, France, Portugal, Russia, Ukraine</td>
</tr>
<tr>
<td>Doleron</td>
<td>Slovenia</td>
</tr>
<tr>
<td>Depon</td>
<td>Greece</td>
</tr>
<tr>
<td>Doxamol</td>
<td>Israel</td>
</tr>
<tr>
<td>Dolex</td>
<td>Colombia</td>
</tr>
<tr>
<td>Doliprane</td>
<td>France, Portugal, Russia, Ukraine</td>
</tr>
<tr>
<td>Efferalgan</td>
<td>France, Italy, Portugal, Russia, Spain, Ukraine</td>
</tr>
<tr>
<td>FeverAll</td>
<td>United States</td>
</tr>
<tr>
<td>Gelocateil</td>
<td>Spain</td>
</tr>
<tr>
<td>Grinp</td>
<td>Turkey</td>
</tr>
<tr>
<td>Lekadol</td>
<td>Croatia, Slovenia</td>
</tr>
<tr>
<td>Metacin</td>
<td>India</td>
</tr>
<tr>
<td>Panol</td>
<td>Denmark, Finland, France</td>
</tr>
<tr>
<td>Panado</td>
<td>South Africa</td>
</tr>
<tr>
<td>Panadol</td>
<td>Australia, Azerbaijan, Central America, Egypt, Finland, Greece, Hong Kong, Hungary, Indonesia, Ireland, Kenya, Lebanon, Macedonia, Malaysia, Malta, Netherlands, New Zealand, Nigeria, Pakistan, Poland, Portugal, Romania, Russia, Saudi Arabia, Singapore, Sri Lanka, Switzerland, Taiwan, Ukraine, Estonia, United Kingdom</td>
</tr>
<tr>
<td>Panamax</td>
<td>Australia, United Kingdom</td>
</tr>
<tr>
<td>Panodil</td>
<td>Denmark, Iceland, Sweden</td>
</tr>
<tr>
<td>Paracet</td>
<td>Norway</td>
</tr>
<tr>
<td>Paralen</td>
<td>Czech Republic, Slovakia</td>
</tr>
<tr>
<td>Paramed</td>
<td>Botswana, South Africa, Zimbabwe</td>
</tr>
<tr>
<td>Paramol</td>
<td>Israel, Taiwan</td>
</tr>
<tr>
<td>Perdolan</td>
<td>Belgium</td>
</tr>
<tr>
<td>Perfalgan</td>
<td>Germany</td>
</tr>
<tr>
<td>Pinex</td>
<td>Denmark, Iceland, Norway</td>
</tr>
<tr>
<td>Plicet</td>
<td>Croatia</td>
</tr>
<tr>
<td>Reliv</td>
<td>Sweden</td>
</tr>
<tr>
<td>Rokamol</td>
<td>Israel</td>
</tr>
<tr>
<td>Sara</td>
<td>Thailand</td>
</tr>
<tr>
<td>Tachiprina</td>
<td>Italy</td>
</tr>
<tr>
<td>Tylenol</td>
<td>Brazil, Canada, Japan, South Korea, Thailand, United States</td>
</tr>
<tr>
<td>Tempra</td>
<td>Philippines</td>
</tr>
</tbody>
</table>
### Entrez Gene

**DMD**

<table>
<thead>
<tr>
<th>Official Symbol</th>
<th>and Name: dystrophin (muscular dystrophy, Duchenne and Becker types) [Homo sapiens]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OtherAliases:</td>
<td>GS1-19O24.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272</td>
</tr>
<tr>
<td>OtherDesignations:</td>
<td>Duchenne muscular dystrophy protein: dystrophin</td>
</tr>
<tr>
<td>Chromosome:</td>
<td>X, Location: Xp21.2</td>
</tr>
<tr>
<td>Annotation:</td>
<td>Chromosome X, NC_000023.9 (33267646..31047265, complement)</td>
</tr>
<tr>
<td>MIM:</td>
<td>30037</td>
</tr>
<tr>
<td>GeneID:</td>
<td>1756</td>
</tr>
</tbody>
</table>
Names for renal cell carcinoma

http://www.clininfo.co.uk/clue5/clue.htm
Entity recognition

- Identifying biomedical entities in text
  - Names entity recognition
  - Tagging “mentions”
  - Semantic annotation

- Supported by terminology
  - Collects the names used in the domain
  - Often incompletely

- Example: BioCreative
  - 1A – Gene name identification
  - 2GM – Gene mention tagging
Normalization

- Biomedical entities are identified by unique identifiers in various terminology systems
- Resolve names into identifiers (in a given namespace)
- Supported (in part) by terminology resources
- Example: BioCreative
  - 1B and 2GN – Gene Normalization

Critical Assessment for Information Extraction in Biology
## Identifier for paracetamol (acetaminophen)

<table>
<thead>
<tr>
<th>Source</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master Drug Data Base. Medi-Span</td>
<td>5005</td>
<td>Acetaminophen</td>
</tr>
<tr>
<td>FDA National Drug Code Directory</td>
<td>50612</td>
<td>PARACETAMOL</td>
</tr>
<tr>
<td>FDA Structured Product Labels</td>
<td>362O9ITL9D</td>
<td>ACETAMINOPHEN</td>
</tr>
<tr>
<td>First DataBank NDDF Plus</td>
<td>001605</td>
<td>Acetaminophen</td>
</tr>
<tr>
<td>SNOMED Clinical Terms</td>
<td>90332006</td>
<td>Acetaminophen (product)</td>
</tr>
<tr>
<td>SNOMED Clinical Terms</td>
<td>387517004</td>
<td>Acetaminophen (substance)</td>
</tr>
<tr>
<td>VA National Drug File</td>
<td>4017513</td>
<td>ACETAMINOPHEN</td>
</tr>
</tbody>
</table>

Source: RxNorm database (5/3/2007)
**Identifier for dystrophin**


<table>
<thead>
<tr>
<th>DMD</th>
<th>Order cDNA clone, Links</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Official Symbol</strong> DMD and Name: dystrophin (muscular dystrophy, Duchenne and Becker types) [Homo sapiens]</td>
<td></td>
</tr>
<tr>
<td><strong>Other Aliases:</strong> GS1-19O24.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272</td>
<td></td>
</tr>
<tr>
<td><strong>Other Designations:</strong> Duchenne muscular dystrophy protein: dystrophin</td>
<td></td>
</tr>
<tr>
<td><strong>Chromosome:</strong> X, <strong>Location:</strong> Xp21.2</td>
<td></td>
</tr>
<tr>
<td><strong>Annotation:</strong> Chromosome X, NC_000023.9 (33267646..31047265, complement)</td>
<td></td>
</tr>
<tr>
<td><strong>MIM:</strong> 300377</td>
<td></td>
</tr>
<tr>
<td><strong>GeneID:</strong> 1756</td>
<td></td>
</tr>
</tbody>
</table>
Identifier for renal cell carcinoma

http://www.clininfo.co.uk/clue5/clue.htm
Mapping / Integration

- Identify equivalent entities across systems (across namespaces)
  - Shared identifiers
  - Existing mappings (e.g., SNOMED CT to ICD-9-CM)
  - Ontology alignment techniques (lexical + structural)

- Align equivalent entities
  - Pairwise: mapping
  - More broadly: integration

- Forms the basis for information integration in the Semantic Web (mashups)
## Identifier for paracetamol (acetaminophen)

<table>
<thead>
<tr>
<th>Source</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master Drug Data Base. Medi-Span</td>
<td>5005</td>
<td>Acetaminophen</td>
</tr>
<tr>
<td>FDA National Drug Code Directory</td>
<td>50612</td>
<td>PARACETAMOL</td>
</tr>
<tr>
<td>FDA Structured Product Labels</td>
<td>362O9ITL9D</td>
<td>ACETAMINOPHEN</td>
</tr>
<tr>
<td>First DataBank NDDF Plus</td>
<td>001605</td>
<td>Acetaminophen</td>
</tr>
<tr>
<td>SNOMED Clinical Terms</td>
<td>90332006</td>
<td>Acetaminophen (product)</td>
</tr>
<tr>
<td>SNOMED Clinical Terms</td>
<td>387517004</td>
<td>Acetaminophen (substance)</td>
</tr>
<tr>
<td>VA National Drug File</td>
<td>4017513</td>
<td>ACETAMINOPHEN</td>
</tr>
<tr>
<td>RxNorm</td>
<td>161</td>
<td>Acetaminophen</td>
</tr>
</tbody>
</table>
**Identifier for dystrophin**


<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DMD</strong></td>
<td>Official Symbol DMD and Name: dystrophin (muscular dystrophy, Duchenne and Becker types) [Homo sapiens]</td>
</tr>
<tr>
<td></td>
<td>Other Aliases: GS1-19O24.1, BMD, CMD3B, DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272</td>
</tr>
<tr>
<td></td>
<td>Other Designations: Duchenne muscular dystrophy protein: dystrophin</td>
</tr>
<tr>
<td></td>
<td>Chromosome: X, Location: Xp21.2</td>
</tr>
<tr>
<td></td>
<td>Annotation: Chromosome X, NC_000023.9 (33267646..31047265, complement)</td>
</tr>
<tr>
<td></td>
<td>MIM: 300377</td>
</tr>
<tr>
<td></td>
<td>GeneID: 1756</td>
</tr>
</tbody>
</table>
Identifier for renal cell carcinoma

http://www.clininfo.co.uk/clue5/clue.htm
Information integration in biomedicine

Semantic Web perspective
HCLS mashup

432508
Entrez-Gene 432508
cleavage and polyadenylation specific factor 6
location: nucleus

Transcript Region
NC_000076.4

Genomic Context

Open Map View

http://esw.w3.org/topic/HCLS/HCLSIG_DemoHomePage_HCLSIG_Demo
From *glycosyltransferase* to *congenital muscular dystrophy*
Terminology integration in biomedicine

Unified Medical Language System
Motivation

- Started in 1986
- National Library of Medicine
- “Long-term R&D project”

“[…] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information. The first is **the variety of ways the same concepts are expressed** in different machine-readable sources and by different people. The second is the **distribution** of useful information among many disparate databases and systems.”
Unified Medical Language System

- SPECIALIST Lexicon
  - 200,000 lexical items
  - Part of speech and variant information

- Metathesaurus
  - 5M names from over 100 terminologies
  - 1M concepts
  - 16M relations

- Semantic Network
  - 135 high-level categories
  - 7000 relations among them
Addison’s disease

Example
Addison’s disease in medical vocabularies

**Synonyms**
- Addisonian syndrome
- Bronzed disease
- Addison melanoderma
- Asthenia pigmentosa
- Primary adrenal deficiency
- Primary adrenal insufficiency
- Primary adrenocortical insufficiency
- Chronic adrenocortical insufficiency

- eponym
- symptoms
- clinical
- variants
Organize terms

- Synonymous terms clustered into a concept
- Preferred term
- Unique identifier (CUI)

<table>
<thead>
<tr>
<th>Term</th>
<th>MeSH</th>
<th>D000224</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison Disease</td>
<td>MeSH</td>
<td>D000224</td>
</tr>
<tr>
<td>Primary hypoadrenalism</td>
<td>MedDRA</td>
<td>10036696</td>
</tr>
<tr>
<td>Primary adrenocortical insufficiency</td>
<td>ICD-10</td>
<td>E27.1</td>
</tr>
<tr>
<td>Addison's disease (disorder)</td>
<td>SNOMED CT</td>
<td>363732003</td>
</tr>
<tr>
<td>C0001403</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Addison's disease
Metathesaurus Concepts (2007AA)

- **Concept** (~ 1.4 M) CUI
  - Set of synonymous concept names
- **Term** (~ 4.9 M) LUI
  - Set of normalized names
- **String** (~ 5.5 M) SUI
  - Distinct concept name
- **Atom** (~ 6.8 M) AUI
  - Concept name in a given source

<table>
<thead>
<tr>
<th>CUI</th>
<th>Term</th>
<th>String</th>
<th>Atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0000001</td>
<td>headache (source 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000002</td>
<td>headache (source 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0000001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000003</td>
<td>Headache (source 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000004</td>
<td>Headache (source 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0000002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0000001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0000005</td>
<td>Cephalgia (source 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S0000003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0000002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0000001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Addison’s Disease: Concept

A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.
**Metathesaurus** Evolution over time

- **Concepts never die (in principle)**
  - CUIs are permanent identifiers

- **What happens when they do die (in reality)?**
  - Concepts can merge or split
  - Resulting in new concepts and deletions

---

**Addison's disease**

- [C0001403]

**Addison's disease, NOS**

- [C0271735]
Diseases of the endocrine system

Diseases of the Adrenal Glands

Addison’s Disease
Endocrine Diseases

Adrenal Gland Diseases

Adrenal Gland Hypofunction

Addison’s Disease
Endocrine disorder

Adrenal disorder

Adrenal cortical disorder

Adrenal cortical hypofunction

Addison’s Disease
Endocrine disorder

Disorder of adrenal gland

Hypoadrenalism

Adrenal Hypofunction

Corticoadrenal insufficiency

Addison’s Disease
Primary adrenocortical insufficiency

Other disorders of adrenal gland

Disorders of other endocrine gland
Organize concepts

- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)
Adrenal Cortex Diseases

- Hypoadrenalism
  - Adrenal Gland Hypofunction
    - Adrenal cortical hypofunction
      - Addison’s Disease

Endocrine Diseases

SNOMED
MeSH
AOD
Read Codes

UMLS
Endocrine System

Endocrine Glands

Abdominal organ

Diseases

Endocrine Diseases

Adrenal Dysfunction

Adrenal Gland Diseases

Adrenal Cortex Diseases

Adrenal Cortex Dysfunction

Hypoadrenalism

Adrenal Gland Hypofunction

Adrenal cortical hypofunction

Secondary hypocortisolism

Addison’s Disease

Addison’s disease due to autoimmunity

Disorders of other endocrine gland

Other disorders of adrenal gland
Source Vocabularies

- 139 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.5M names
  - 1.4M concepts
  - 16M relations
- Common presentation
Biomedical terminologies

◆ General vocabularies
  - anatomy (UWDA, Neuronames)
  - drugs (RxNorm, First DataBank, Micromedex, …)
  - medical devices (UMD, SPN)

◆ Several perspectives
  - clinical terms (SNOMED CT)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - data exchange terminologies (HL7, LOINC)
Biomedical terminologies (cont’d)

◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (NCI Thesaurus, PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART, MedDRA)
  - primary care (ICPC)
  - genomics (Gene Ontology, HUGO, OMIM)

◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Other subdomains
- MeSH
- SNOMED
- OMIM
- Genome annotations
- Model organisms
- NCBI Taxonomy
- UWDA
- GO
- Anatomy
- ...
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains
How do they do that?

- Lexical knowledge
- Semantic pre-processing
- UMLS editors
Lexical knowledge

- Adrenal gland diseases
- Adrenal disorder
- Disorder of adrenal gland
- Diseases of the **adrenal glands**
- C0001621
Semantic pre-processing

- Metadata in the source vocabularies
- Tentative categorization
- Positive (or negative) evidence for tentative synonymy relations based on lexical features
UMLS vs. Semantic Web
Similarities, differences and unresolved issues

- Identifying biomedical entities
  - Trans-namespace integration
  - No UMLS-based URIs
- Availability
  - Intellectual property restrictions
  - Application Programming Interface
- Formats
  - RRF vs. SW languages
- UMLS as an ontology?
  - Underspecified semantics
Identifying biomedical entities

- Syntax vs. semantics
  - URI, LSID, ... vs. reference ontologies

- Integrative resources vs. individual namespaces
  - Unified Medical Language System (UMLS) vs. GO, MeSH, SNOMED, ...
No UMLS-based URIs  Syntax

- No officially supported UMLS-based URIs for biomedical entities
  e.g., http://umls.org/C0001403

- Possible alternatives
  - Redirection service (e.g., PURL) http://purl.org/

- Resolution issues: what is expected to be returned?
  - Acknowledgment of existence
  - Preferred term
  - Set of names, relations,… in RDF
No UMLS-based URIs  Semantics

- Potential resources for trans-namespace identification of biomedical entities
  - Clinical medicine: UMLS CUIs
  - [Genomics: Entrez Gene]

- Ontology of biomedical relationships
  - No comprehensive integrative resource available
    - OBO relations
    - UMLS Semantic Network relations
    - GALEN relations
Trans-namespace integration

Addison's disease (disorder) (363732003)

Other subdomains

Model organisms

NCBI Taxonomy

UWDA

UMLS C0001403

Clinical repositories

SNOMED

OMIM

MeSH

Biomedical literature

Addison Disease (D000224)

Genetic knowledge bases

Genome annotations

Anatomy

GO

MeSH

Biomedical literature
Trans-namespace integration

◆ Advantages

- Over shared identifiers (increased recall)
- Over lexical mapping (increased recall + precision)

<table>
<thead>
<tr>
<th>Addison Disease: MeSH:D000224</th>
<th>X</th>
<th>Primary adrenocortical insufficiency: ICD9CM:E27.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>UMLS:C0001403</td>
</tr>
</tbody>
</table>
Ambiguity resolution

NF2

- Neurofibromatosis 2 [disease]
  C0027832

- Neurofibromin 2 [protein]
  C0254123

- Neurofibromatosis 2 gene [gene]
  C0085114
Availability

Intellectual property restrictions

- UMLS: free license required

- Some intellectual property restrictions
  - 2/3 of the names freely available (in the US)

<table>
<thead>
<tr>
<th>SRL</th>
<th>Source Count</th>
<th>% of Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2181959</td>
<td>32.22%</td>
</tr>
<tr>
<td>1</td>
<td>94059</td>
<td>1.39%</td>
</tr>
<tr>
<td>2</td>
<td>22156</td>
<td>0.33%</td>
</tr>
<tr>
<td>3</td>
<td>2111546</td>
<td>31.18%</td>
</tr>
<tr>
<td>4</td>
<td>2362949</td>
<td>34.09%</td>
</tr>
<tr>
<td>0+4</td>
<td>4544908</td>
<td>67.11%</td>
</tr>
</tbody>
</table>

- Web browser: username/password required

http://www.nlm.nih.gov/research/umls/
### Availability Application Programming Interfaces

- Remote server at NLM
- Local application connected through

<table>
<thead>
<tr>
<th>Java RMI</th>
<th>TCP/IP socket</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Java-based applications</td>
<td>- XML-based queries</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>- XML schema</td>
</tr>
<tr>
<td>- Set of Java classes</td>
<td>- Socket server</td>
</tr>
<tr>
<td>(part of the UMLSKS API</td>
<td>- Host: umlsks.nlm.nih.gov</td>
</tr>
<tr>
<td>download)</td>
<td>- Port: 8042</td>
</tr>
<tr>
<td>- Detailed Javadoc</td>
<td></td>
</tr>
<tr>
<td>documentation online and</td>
<td></td>
</tr>
<tr>
<td>with API download</td>
<td></td>
</tr>
</tbody>
</table>
Availability Web Services-based API

- Part of the Knowledge Source Server version 3
  - Portlet-based, customizable
  - WS architecture
- Coming soon
  - Alpha release in July 2007
  - Beta release in November 2007
Representation formalism

**UMLS**
- Rich Release Format (RRF)
- [Original Release Format (ORF)]
- Support for source transparency

**Semantic Web**
- RDF – Resource Description Framework
- OWL – Web Ontology Language
- SKOS – Simple Knowledge Organization Systems

**Other formats**
- OBO – Open Biological Ontologies [http://obo.sourceforge.net/browse.html](http://obo.sourceforge.net/browse.html)
- LexGrid [http://informatics.mayo.edu/LexGrid/](http://informatics.mayo.edu/LexGrid/)

**Converters**
UMLS vocabularies available in RDF/OWL

- **NCI Thesaurus (OWL)**
  - [http://ncicb.nci.nih.gov/core/EVS](http://ncicb.nci.nih.gov/core/EVS)

- **Gene Ontology**

- **Repository of biomedical ontologies (OBO, OWL)**
Porting vocabularies to OWL  

Experiments

- **MeSH**
  - Soualmia et al., KR-MED 2004

- **Foundational Model of Anatomy (FMA)**
  - Golbreich et al., JWS 2006 (OWL DL)
  - Noy and Rubin, SMI Tech Report 2007 (OWL Full)

- **UMLS Semantic Network**
  - Kashyap and Borgida, ISWC 2003

- **UMLS Metathesaurus**
  - Cornet and Abu-Hanna, AMIA 2002
Neurofibromatosis 2
(Type II neurofibromatosis, Bilateral acoustic neurofibromatosis)
C0027832

NF2
(Neurofibromin 2 gene)  
C0085114

Merlin
(Schwannomin, Neurofibromin 2)  
C0254123

Merlin, Drosophila

Drosophila melanogaster merlin (Dmerlin) mRNA, complete cds.  
U49724

UMLS Metathesaurus (Concepts and relations)

Gene or Genome

Tumor suppressor genes

Biologically Active Substance

Tumor suppressor proteins

Amino Acid, Peptide, or Protein

Neoplastic Process

Benign neoplasms of cranial nerves

Neurofibromatoses

UMLS as an “ontology”
Limitations

- Genes not systematically represented
  - Most gene products and diseases are
- Gene/Gene product-Disease relations
  - Not systematically represented
  - Not explicitly represented (e.g., co-occurrence)
- Cross-references not systematically represented
- Naming conventions (genes)
Underspecified semantics

- Relationship “attribute” not always present
- Relations used to create hierarchies vs. hierarchical relations

Environment and Public Health [G03]
Public Health [G03.850]
Accidents [G03.850.110]

Accident Prevention [G03.850.110.060] +
Accidental Falls [G03.850.110.085]
Accidents, Aviation [G03.850.110.185]
Accidents, Home [G03.850.110.205]
Accidents, Occupational [G03.850.110.250] +
Accidents, Radiation [G03.850.110.285]
Accidents, Traffic [G03.850.110.320]
Drowning [G03.850.110.500] +
Summary
Biomedicine and Semantic Web

- Semantic Web technologies have not been widely adopted yet in biomedicine
  - OBO vs. OWL
  - caBIG vs. Taverna
- Use cases
  - Information/Data integration
- Recent efforts
  - W3C Health Care and Life Sciences Interest Group
UMLS and Semantic Web

- Terminology integration
- Based on existing terminologies
- Trans-namespace, permanent identifiers
- APIs available
  - Web Services-based API coming soon
- Can support information integration
- “Proprietary” representation (RRF)
- Some intellectual property restrictions
- Underspecified semantics
- No UMLS-based URIs
Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
UMLS References

- **UMLS**
  [umlsinfo.nlm.nih.gov](umlsinfo.nlm.nih.gov)

- **UMLS browsers**
  (free, but UMLS license required)
  
  
  
  - RRF browser
    (standalone application distributed with the UMLS)
UMLS References

Gentle introduction


Seminal paper

Semantic Web for Health Care and Life Sciences

◆ W3C Health Care and Life Sciences Interest Group
  ● http://www.w3.org/2001/sw/hcls/


◆ Demo presented at the WWW2007 conference (May 2007)
  http://esw.w3.org/topic/HCLS/HCLSIG_DemoHomePage_HCLSIG_Demo
Biomedical information integration through RDF

◆ Biomedical perspective

  

◆ Semantic Web perspective

  