Disclaimer

The views and opinions expressed do not necessarily state or reflect those of the U.S. Government, and they may not be used for advertising or product endorsement purposes.
Overview

- An example
- Types of resources for mining biomedical text
- Three types of resources
  - Lexical resources – SPECIALIST Lexicon, norm
  - Terminological resources – Metathesaurus, MetaMap
  - Ontological resources – Semantic Network, SemRep
An example

Neurofibromatosis 2
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.

Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.

- vestibular schwannomas *manifestation of* neurofibromatosis 2
- neurofibromatosis 2 *associated with* mutation of NF2 gene
- NF2 gene *located on* chromosome 22
Types of resources for mining biomedical text
Types of resources

◆ Lexical resources
  ● Collections of lexical items
  ● Additional information
    ■ Part of speech
    ■ Spelling variants
  ● Useful for entity recognition
  ● UMLS SPECIALIST Lexicon, WordNet

◆ Ontological resources
  ● Collections of
    ■ kinds of entities (substances, qualities, processes)
    ■ relations among them
  ● Useful for relation extraction
  ● UMLS Semantic Network, BioTop

◆ Terminological resources
  ● Collections lexical items + identifiers
  ● Useful for entity resolution
  ● UMLS Metathesaurus
Types of resources (revisited)

◆ Lexical and terminological resources
  ● Mostly collections of names for biomedical entities
  ● Often have some kind or hierarchical organization (e.g., relations)

◆ Ontological resources
  ● Mostly collections of relations among biomedical entities
  ● Sometimes also collect names
Lexical / Ontological MeSH


- Endocrine system diseases
  - Adrenal gland diseases
    - Adrenal Insufficiency
      - Addison Disease
  - Immune system diseases
    - Autoimmune diseases

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Entry Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addison Disease</td>
<td>Addison's Disease</td>
</tr>
<tr>
<td></td>
<td>Primary Adrenal Insufficiency</td>
</tr>
<tr>
<td></td>
<td>Primary Adrenocortical Insufficiency</td>
</tr>
<tr>
<td></td>
<td>Primary Hypoadrenalism</td>
</tr>
</tbody>
</table>
Lexical / Ontological FMA

Foundational Model Explorer

Select navigation tree type: subclass

PREFERRED NAME:
Heart

NON-ENGLISH EQUIVALENT:

<table>
<thead>
<tr>
<th>name</th>
<th>language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cor</td>
<td>Latin</td>
</tr>
<tr>
<td>Corazon</td>
<td>Spanish</td>
</tr>
<tr>
<td>Coeur</td>
<td>French</td>
</tr>
<tr>
<td>Herz</td>
<td>German</td>
</tr>
<tr>
<td>Cuore</td>
<td>Italian</td>
</tr>
<tr>
<td>Puso</td>
<td>Filipino</td>
</tr>
</tbody>
</table>

FMAID:
7088

DEFINITION:
Organ with cavitated organ parts, which is continuous with the systemic and pulmonary arterial and venous trees. Examples: There is only one heart.
Unified Medical Language System

◆ SPECIALIST Lexicon
  ● 650,000 lexical items
  ● Part of speech and variant information

◆ Metathesaurus
  ● ~10M names from over 100 terminologies
  ● ~3M concepts
  ● >10M relations

◆ Semantic Network
  ● 133 high-level categories
  ● 7000 relations among them

Lexical resources
- LVG / Norm
- MetaMap

Terminological resources
- SemRep

Ontological resources
Lexical resources

SPECIALIST Lexicon
and lexical tools

Lexical Systems Group

http://umlslex.nlm.nih.gov/
SPECIALIST Lexicon

◆ Content
  ● English lexicon
  ● Many words from the biomedical domain

◆ 650,000 lexical items

◆ Word properties
  ● morphology
  ● orthography
  ● syntax

◆ Used by the lexical tools
Morphology

◆ Inflection
  - noun: nucleus, nuclei
  - verb: cauterize, cauterizes, cauterized, cauterizing
  - adjective: red, redder, reddest

◆ Derivation
  - verb ↔ noun: cauterize -- cauterization
  - adjective ↔ noun: red -- redness
Orthography

◆ Spelling variants

- oe/e  
  oesophagus - esophagus

- ae/e  
  anaemia - anemia

- ise/ize  
  cauterise - cauterize

- genitive mark  
  Addison's disease
  Addison disease
  Addisons disease
Syntax

◆ Complementation

  ● verbs
    - intransitive: I'll treat.
    - transitive: He treated the patient.
    - ditransitive: He treated the patient with a drug.

  ● nouns
    - prepositional phrase: Valve of coronary sinus

◆ Position for adjectives
SPECIALIST Lexicon record

{
  base=hemoglobin (base form)
  spelling_variant=haemoglobin
  entry=E0031208 (identifier)
  cat=noun (part of speech)
  variants=uncount (no plural)
  variants=reg (plural: hemoglobins, hemoglobins)
}
Lexical tools

- To manage lexical variation in biomedical terminologies
- Major tools
  - Normalization
  - Indexes
  - Lexical Variant Generation program (lvg)
- Based on the SPECIALIST Lexicon
- Used by noun phrase extractors, search engines
Normalization

Remove genitive

Hodgkin’s diseases, NOS

Hodgkin diseases, NOS

Remove stop words

Hodgkin diseases,

Hodgkin diseases,

Lowercase

hodgkin diseases,

hodgkin diseases,

Strip punctuation

hodgkin diseases

hodgkin diseases

Uninflect

hodgkin disease

hodgkin disease

Sort words

disease hodgkin

disease hodgkin

Lister Hill National Center for Biomedical Communications
Normalization: Example

- Hodgkin Disease
- HODGKINS DISEASE
- Hodgkin's Disease
- Disease, Hodgkin's
- Hodgkin's, disease
- HODGKIN'S DISEASE
- Hodgkin's disease
- Hodgkins Disease
- Hodgkin's disease NOS
- Hodgkin's disease, NOS
- Disease, Hodgkins
- Diseases, Hodgkins
- Hodgkins Diseases
- Hodgkins disease
- hodgkin's disease
- Disease, Hodgkin

normalize  
disease hodgkin
Normalization Applications

- Model for lexical resemblance
- Help find lexical variants for a term
  - Terms that normalize the same usually share the same LUI
- Help find candidates to synonymy among terms
- Help map input terms to UMLS concepts
Indexes

◆ Word index
  ● word to Metathesaurus strings
  ● one word index per language

◆ Normalized word index
  ● normalized word to Metathesaurus strings
  ● English only

◆ Normalized string index
  ● normalized term to Metathesaurus strings
  ● English only
Lexical Variant Generation program

- Tool for specialists (linguists)
- Performs atomic lexical transformations
  - generating inflectional variants
  - lowercase
  - ...
- Performs sequences of atomic transformations
  - a specialized sequence of transformations provides the normalized form of a term (the norm program)
Related NLM tools

The SPECIALIST Text Tools includes tokenizers that analyze text into word, term, phrase, sentence and section pieces. The tools also include a variant lookup module that retrieves variant ways of expressing the phrases found in the text. The tools are intended to analyze documents into instances of document objects.

The tools are written in Java. These tools include the following:

- a word/Sentence/section Tokenizer
- a term tokenizer
- a phrase tokenizer
- a term variant lookup
- a part-of-speech tagger (client)
- a document index maker
- a tool to create the textTool indexes

The SPECIALIST spelling resources include two programs GSpell, a spelling suggestion tool and BagOwordsPlus a phrase retrieval tool.

GSpell uses several word similarity algorithms to suggest correct spellings for misspelled words. Unlike other spelling suggestion programs GSpell treats space as it would any other letter so that GSpell can correct errors in word compounding. GSpell also be used in word similarity tasks that do not involve misspelling.

BagOWordsPlus uses the word similarity algorithms of GSpell to perform word similarity based phrase level information retrieval.

The dTagger is a Part of Speech (POS) tagger. A POS tagger assigns part of speech tags such as noun, adjective, adverb to sentences. Such tag assignments are a needed component to determining phrase boundaries and head assignment. The dTagger includes the following features: It can tokenize text into single or multi-word terms. It is built specifically for use with the SPECIALIST Lexicon. A default trained model is included, trained on a set of annotated MEDLINE abstracts in the genomics field, (the MedPost corpus). The trainer and updater programs are included to allow the creation of new trained models. Models can be updated with lots of untagged text. Can be trained with just untagged text, if need be. The dTagger is an open source resource and is freely available subject to these terms and conditions.

http://umlslex.nlm.nih.gov/
Lexical resources

Other resources
Need for additional resources

- More generic
  - WordNet

- More specific
  - Lexical items specific to specialized subdomains
    - Not listed in biolexicons
    - Not amenable to normalization
  - Examples
    - Genes, proteins
      - MAPK3 / Mapk3 / mapk3
    - Chemicals
      - 5’-3’ exonuclease / 3’-5’ exonuclease
    - Drugs
    - Acronyms
Gene and protein names

◆ Additional resources

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HGNC</td>
<td><a href="http://www.genenames.org/">http://www.genenames.org/</a></td>
</tr>
<tr>
<td>UniProt</td>
<td><a href="http://www.ebi.uniprot.org/index.shtml">http://www.ebi.uniprot.org/index.shtml</a></td>
</tr>
</tbody>
</table>

◆ Additional identification methods

- e.g., ABGene (Tanabe & Wilbur, NCBI)
- BioCreAtIvE
  - Gene mention identification
  - Gene normalization

[Additional resources image]
Chemical names

◆ Additional resources

<table>
<thead>
<tr>
<th>Additional resources</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChEBI</td>
<td><a href="http://www.ebi.ac.uk/chebi/">http://www.ebi.ac.uk/chebi/</a></td>
</tr>
</tbody>
</table>
Drug names

- Covered by UMLS
- Specialized resource: RxNorm
  - Branded names / generic names
  - Various levels of aggregation
    - Ingredient
    - Ingredient + dose
    - Ingredient + form
    - Ingredient + dose + form
  - Codes in various reference systems
- Mostly US drugs, no “over-the-counter” drugs
Acronyms

◆ Many resources available

- AcroMine
  http://www.nactem.ac.uk/software/acromine/
- Stanford Biomedical Abbreviation Server
  http://www.medlingmap.org/taxonomy/term/120
- AcroMed
  http://www.medstract.org/
- SaRAD
Terminological resources

UMLS Metathesaurus

http://www.nlm.nih.gov/research/umls/
Source Vocabularies

- 139 families of source vocabularies
  - Not counting translations
- 21 languages
- Broad coverage of biomedicine
  - 8.6M names (normalized)
  - ~3M concepts
  - >10M relations
- Common presentation
Biomedical terminologies

- **General vocabularies**
  - anatomy (FMA, Neuronames)
  - drugs (RxNorm, ATC, First DataBank, Micromedex)
  - medical devices (UMD, SPN)

- **Several perspectives**
  - clinical terms (SNOMED CT)
  - information sciences (MeSH)
  - administrative terminologies (ICD-9/10-CM, CPT-4)
  - data exchange terminologies (HL7, LOINC)
Biomedical terminologies (cont’d)

◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, ICNP)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (MedDRA, WHO ART)
  - primary care (ICPC)

◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)
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>
</tbody>
</table>

C0001403

Addison's disease
Organize concepts

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

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

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Trans-namespace resolution (1)

Neurofibromatosis, type 2 (92503002)

Clinical repositories

Genetic knowledge bases

UMLS C0027832

Biomedical literature

Neurofibromatosis 2 (D016518)

Genome annotations

Model organisms

NCBI Taxonomy

FMA

GO

Anatomy

SNOMED CT

OMIM

Other subdomains
Trans-namespace resolution (2)

Source: Multum [generic drug]

Nizoral, 200 mg oral tablet (MMSL:2140)

Ketoconazole 200 MG Oral Tablet [Nizoral] (RxNorm:201896)

Ketoconazole 200 MG Oral Tablet (RxNorm:197853)

Ketoconazole Tab 200 MG (MDDB:13317)

Nizoral (RxNorm:202692)

Ketoconazole (RxNorm:6135)

tradename of

has ingredient

tradename of

Target: Medi-Span [generic drug]

Source: Multum [generic drug]

Nizoral, 200 mg oral tablet (MMSL:2140)

Ketoconazole 200 MG Oral Tablet [Nizoral] (RxNorm:201896)

Ketoconazole 200 MG Oral Tablet (RxNorm:197853)

Ketoconazole Tab 200 MG (MDDB:13317)

Nizoral (RxNorm:202692)

Ketoconazole (RxNorm:6135)

tradename of

has ingredient

tradename of

Target: Medi-Span [generic drug]

Source: Multum [generic drug]

Nizoral, 200 mg oral tablet (MMSL:2140)

Ketoconazole 200 MG Oral Tablet [Nizoral] (RxNorm:201896)

Ketoconazole 200 MG Oral Tablet (RxNorm:197853)

Ketoconazole Tab 200 MG (MDDB:13317)

Nizoral (RxNorm:202692)

Ketoconazole (RxNorm:6135)

tradename of

has ingredient

tradename of

Target: Medi-Span [generic drug]
Terminological resources

MetaMap

Indexing Initiative
http://ii.nlm.nih.gov/
MetaMap

- UMLS-based entity recognition system
  - Linguistically motivated
  - Exploits both the SPECIALIST lexicon and Metathesaurus

- In practice, used to identify UMLS concepts in biomedical text

- Freely available (UMLS license)

- Two versions
  - Web-based
  - Standalone (MMTx)
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Terminological resources

Other resources
Other NER systems  TerMine

TerMine (C-value) analysis

Found 5 terms in 2.2 seconds - all terms (in table) (in text) - threshold: 0 Apply

Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.

Thank you for using TerMine. Please now complete a questionnaire to let us know your views about this service.
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Resulting tagged text

Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Other NER systems  NCBO Annotator

Annotator
Get annotations for biomedical text with concepts from the ontologies

Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.

Select Ontologies
Type here to select ontologies or leave blank to use all

Select UMLS Semantic Types
Type here to select UMLS semantic types

- Match longest only
- Include Manual Mappings
- Include Ancestors Up To Level: None

Get Annotations
<table>
<thead>
<tr>
<th>CLASS</th>
<th>ONTOLOGY</th>
<th>TYPE</th>
<th>CONTEXT</th>
<th>MATCHED CLASS</th>
<th>MATCHED ONTOLOGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>neurofibromatosis type 2 (NF2)</td>
<td>Physician Data Query</td>
<td>direct</td>
<td>Neurofibromatosis type 2 (NF2) is often not ...</td>
<td>neurofibromatosis type 2 (NF2)</td>
<td>Physician Data Query</td>
</tr>
<tr>
<td>Entity</td>
<td>Situation-Based Access Control Ontology</td>
<td>direct</td>
<td>... a distinct entity from peripheral neurofibromatosis. ...</td>
<td>Entity</td>
<td>Situation-Based Access Control Ontology</td>
</tr>
<tr>
<td>Neurofibromatosis Type 2</td>
<td>National Cancer Institute Thesaurus</td>
<td>direct</td>
<td>... peripheral neurofibromatosis. NF2 is a predominantly ...</td>
<td>Neurofibromatosis Type 2</td>
<td>National Cancer Institute Thesaurus</td>
</tr>
<tr>
<td>Neurofibromatosis Type 2</td>
<td>National Cancer Institute Thesaurus</td>
<td>direct</td>
<td>... vestibular schwannomas. NF2 results from a ...</td>
<td>Neurofibromatosis Type 2</td>
<td>National Cancer Institute Thesaurus</td>
</tr>
<tr>
<td>Is a</td>
<td>Systematized Nomenclature of Medicine - Clinical Terms</td>
<td>direct</td>
<td>... neurofibromatosis. NF2 is a predominantly intracranial condition ...</td>
<td>Is a</td>
<td>Systematized Nomenclature of Medicine - Clinical Terms</td>
</tr>
<tr>
<td>Intracranial</td>
<td>Systematized Nomenclature of Medicine - Clinical Terms</td>
<td>direct</td>
<td>... a predominantly intracranial condition whose hallmark ...</td>
<td>Intracranial</td>
<td>Systematized Nomenclature of Medicine - Clinical Terms</td>
</tr>
<tr>
<td>Condition</td>
<td>Neglected Tropical Disease Ontology</td>
<td>direct</td>
<td>... predominantly intracranial condition whose hallmark is ...</td>
<td>Condition</td>
<td>Neglected Tropical Disease Ontology</td>
</tr>
<tr>
<td>BILATERAL</td>
<td>Dermatology Lexicon</td>
<td>direct</td>
<td>... hallmark is bilateral vestibular schwannomas. NF2 ...</td>
<td>BILATERAL</td>
<td>Dermatology Lexicon</td>
</tr>
<tr>
<td>Neuroma, Acoustic</td>
<td>Medical Subject Headings</td>
<td>direct</td>
<td>... is bilateral vestibular schwannomas. NF2 results from ...</td>
<td>Neuroma, Acoustic</td>
<td>Medical Subject Headings</td>
</tr>
<tr>
<td>Results</td>
<td>Ontological Knowledge Base Model for Cystic Fibrosis</td>
<td>direct</td>
<td>... schwannomas. NF2 results from a mutation ...</td>
<td>Results</td>
<td>Ontological Knowledge Base Model for Cystic Fibrosis</td>
</tr>
<tr>
<td>Mutation</td>
<td>Logical Observation Identifier Names and Codes</td>
<td>direct</td>
<td>... from a mutation in the gene ...</td>
<td>Mutation</td>
<td>Logical Observation Identifier Names and Codes</td>
</tr>
<tr>
<td>gene</td>
<td>Single-Nucleotide Polymorphism (SNP) Ontology</td>
<td>direct</td>
<td>... in the gene named merlin, located ...</td>
<td>gene</td>
<td>Single-Nucleotide Polymorphism (SNP) Ontology</td>
</tr>
<tr>
<td>Falco columbarius</td>
<td>Systematized Nomenclature of Medicine, International Version</td>
<td>direct</td>
<td>... gene named merlin, located on chromosome ...</td>
<td>Falco columbarius</td>
<td>Systematized Nomenclature of Medicine, International Version</td>
</tr>
<tr>
<td>Chromosome 22</td>
<td>National Cancer Institute Thesaurus</td>
<td>direct</td>
<td>... located on chromosome 22.</td>
<td>Chromosome 22</td>
<td>National Cancer Institute Thesaurus</td>
</tr>
</tbody>
</table>
Ontological resources
Ontological resources

◆ Provide background knowledge
  ● For resolving ambiguity in entity recognition
    ■ Merlin: Protein or Bird?
  ● For relation extraction
    ■ Template relations between high-level concepts
    ■ Used in combination with clues from linguistic phenomena in text
Ontological resources

◆ Various level of formality
  ● Formal top-level ontologies (e.g., BioTop)
  ● Informal top-level ontologies (e.g., UMLS Semantic Network)
  ● Domain-Range constraints for roles in DL-based terminologies (e.g., SNOMED CT, NCI Thesaurus)
  ● Relations in terminologies

◆ Various level of granularity
  ● UMLS Semantic Network: 133 types
  ● Foundational Model of Anatomy: 70,000 classes
Ontological resources

UMLS Semantic Network
“Biologic Function” hierarchy (isa)

- Biologic Function
  - Physiologic Function
    - Organism Function
      - Mental Process
    - Organ or Tissue Function
    - Cell Function
      - Genetic Function
  - Molecular Function
    - Cell or Molecular Dysfunction
    - Disease or Syndrome
      - Mental or Behavioral Dysfunction
      - Neoplastic Process
  - Pathologic Function
    - Experimental Model of Disease
Associative (non-isa) relationships

- Organism
  - Anatomical Structure
    - Embryonic Structure
    - Anatomical Abnormality
      - Congenital Abnormality
      - Acquired Abnormality
    - Body System
      - Body Part, Organ or Organ Component
      - Tissue
      - Cell
      - Cell Component
      - Gene or Genome
  - Organism Attribute
  - Process of
  - Evaluation of
  - Finding
    - Laboratory or Test Result
    - Sign or Symptom
  - Biologic Function
    - Physiologic Function
    - Pathologic Function
  - Location of
    - Body Location or Region
    - Body Space or Junction
  - Location of
    - Injury or Poisoning
    - Disrupts
    - Co-occurs with
    - Contains, produces
    - Conceptual part of
    - Adjacent to
    - Conceptual part of
Why a semantic network?

- Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently of their position in a hierarchy*.

- A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs.
  - The relationship may or may not hold at the concept level.
  - Other relationships may apply at the concept level.
Relationships *may* inherit semantics.

Semantic Network

- Fully Formed Anatomical Structure
- Body Part, Organ, or Organ Component
- Disease or Syndrome
- Biologic Function
- Pathologic Function
- Adrenal Cortex
- Adrenal Cortical Hypofunction

`isa` indicates a subsumption relationship.

`location of` indicates a location relationship.
Ontological resources

SemRep
Neurofibromatosis type 2 (NF2) is often not recognised as a distinct entity from peripheral neurofibromatosis. NF2 is a predominantly intracranial condition whose hallmark is bilateral vestibular schwannomas. NF2 results from a mutation in the gene named merlin, located on chromosome 22.
Relation extraction “in action”
Ontological resources

Other resources
Other ontological resources

- **Ontologies**
  - Top-level ontologies (e.g., BioTop)
  - Domain ontologies (e.g., FMA, SNOMED CT, NCI Thesaurus, OBO family of ontologies)

- **Many information extraction systems available**
  - Specialized
    - Protein-protein interaction (e.g., Info-PubMed, TextPresso, …)
    - BioCreAtIvE (task 2)
  - More generic (e.g., MedLEE / BioMedLEE)
  - Commercial systems (TeSSI, Linguamatics, …)
Summary
Summary

◆ Lexical and terminological resources enable entity recognition
  ● Terminological resources enable entity resolution

◆ Terminological and ontological resources enable relation extraction

But...

◆ Text mining techniques can also benefit
  ● Specialized lexicons: NER based on machine learning techniques
  ● Terminologies: term extraction / computational terminology
  ● Ontologies: ontology population
References

- Bodenreider O.

  "Lexical, terminological and ontological resources for biological text mining."

UMLS References

◆ Recent overviews


UMLS References

◆ UMLS as a research project


UMLS References

♦ Technical papers


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