Semantic Medline

Note: Access to the Semantic MEDLINE and Semantic MEDLINE Prototype applications and databases requires signing the UMLS Metathesaurus license agreement which I have done.

Story
Slides
Slide 1 Big Data for Government Symposium
Slide 2 Effectively Exploiting Big Data with Semantics: A Pilot Project
Slide 3 Disclaimer
Slide 4 Background 1
Slide 5 Background 2
Slide 6 Challenges
Slide 7 Solution
Slide 8 Pilot project components
Slide 9 The YarcData approach
Slide 10 Semantic processing: SemRep
Slide 11 SemRep: semantic predication
Slide 12 Web application: Semantic MEDLINE
Slide 13 Semantic MEDLINE overview
Slide 14 Use case: Inflammation and cancer
Slide 15 SemMed: semantic predications
Slide 16 Semantic predications as a graph
Slide 17 Cytokine CAUSES Tumorigenesis
Slide 18 MEDLINE citation
Slide 19 Exploiting semantic processing
Slide 20 Acknowledgments
Slide 21 Additional information

Spotfire Dashboard
Spotfire Network Analytics
Semantic MedLine Prototype
Home
Help

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
Summary Tab
Visualization Page
Graph Control Pane
Magnifying Glass
Graph Pane

Figure 1. Colors of High Level Groups for Concepts

Information Pane
Concept Information
Relationship Information
Relation Labels
Search

SEMANTIC RELATION (PREDICATE) DEFINITIONS
ADMINISTERED_TO
AFFECTS
ASSOCIATED_WITH
AUGMENTS
CAUSES
COEXISTS_WITH
CONVERTS_TO
COMPLICATES
DIAGNOSES
DISRUPTS
INHIBITS
INTERACTS_WITH
ISA
LOCATION_OF
MANIFESTATION_OF
METHOD_OF
OCCURS_IN
PART_OF
PRECEDES
PREDISPOSES
PREVENTS
PROCESS_OF
PRODUCES
STIMULATES
TREATS
USES
SEMANTIC TYPE DEFINITIONS (organized by node semantic groups)

Activities and Behaviors
- Activity
- Behavior
- Daily or Recreational Activity
- Event
- Governmental or Regulatory Activity
- Individual Behavior
- Machine Activity
- Occupational Activity
- Social Behavior

Anatomy
- Anatomical Structure
- Body Location or Region
- Body Part, Organ, or Organ Component
- Body Space or Junction
- Body Substance
- Body System
- Cell
- Cell Component
- Embryonic Structure
- Fully Formed Anatomical Structure
- Tissue

Chemicals and Drugs
- Amino Acid, Peptide, or Protein
- Antibiotic
- Biologically Active Substance
- Biomedical or Dental Material
- Carbohydrate
- Chemical
- Chemical Viewed Functionally
- Chemical Viewed Structurally
- Clinical Drug
- Eicosanoid
- Element, Ion, or Isotope
- Enzyme
- Hazardous or Poisonous Substance
- Hormone
Immunologic Factor
Indicator, Reagent, or Diagnostic Aid
Inorganic Chemical
Lipid
Neuroreactive Substance or Biogenic Amine
Nucleic Acid, Nucleoside, or Nucleotide
Organic Chemical
Organophosphorus Compound
Pharmacologic Substance
Receptor
Steroid
Vitamin
Classifications
Classification
Conceptual Entity
Functional Concept
Group Attribute
Idea or Concept
Intellectual Product
Language
Qualitative Concept
Quantitative Concept
Regulation or Law
Spatial Concept
Temporal Concept
Devices
Drug Delivery Device
Medical Device
Research Device
Disorders
Acquired Abnormality
Anatomical Abnormality
Cell or Molecular Dysfunction
Congenital Abnormality
Disease or Syndrome
Experimental Model of Disease
Finding
Injury or Poisoning
Mental or Behavioral Dysfunction
Neoplastic Process
Pathologic Function
Sign or Symptom

Genes and Molecular Sequences
Amino Acid Sequence
Carbohydrate Sequence
Gene or Genome
Molecular Sequence
Nucleotide Sequence

Geographic Areas
Geographic Area

Living Beings
Age Group
Alga
Amphibian
Animal
Archaeon
Bacterium
Bird
Family Group
Fish
Fungus
Group
Human
Invertebrate
Mammal
Organism
Patient or Disabled Group
Plant
Population Group
Professional or Occupational Group
Reptile
Rickettsia or Chlamydia
Vertebrate
Virus

Objects
Entity
Food
Manufactured Object
Physical Object
Substance

Occupations
Biomedical Occupation or Discipline
Occupation or Discipline

Organizations
Health Care Related Organization
Organization
Professional Society
Self-help or Relief Organization

Phenomena
Biologic Function
Environmental Effect of Humans
Human-caused Phenomenon or Process
Laboratory or Test Result
Natural Phenomenon or Process
Phenomenon or Process

Physiology
Cell Function
Clinical Attribute
Genetic Function
Mental Process
Molecular Function
Organ or Tissue Function
Organism Attribute
Organism Function
Physiologic Function

Procedures
Diagnostic Procedure
Educational Activity
Health Care Activity
Laboratory Procedure
Molecular Biology Research Technique
Research Activity
Therapeutic or Preventive Procedure

Download the Predication Database
Description of the Predication Database

Tables

Name: CONCEPT
Name: CONCEPT_SEMTYPE
Name: CONCEPT_TRANSLATION
Name: PREDICATION
Name: PREDICATION_ARGUMENT
Name: SENTENCE
Name: SENTENCE_PREDICATION
Name: PREDICATE_AGGREGATE

Table Relation

Predication Databases

My SQL

Semantic MEDLINE: Semantic Predications for Biomedical Research

Title Page
Introduction
Semantic MEDLINE
SemRep: Extract all predications
Summarize
Visualize as a graph
Link to text 1
Link to text 2
Conduct research
Acknowledgments

Semantic Processing for Managing the Biomedical Research Literature

Title Page
Access to online text 1
Access to online text 2
Emerging applications 1
Emerging applications 2
Automatic semantic interpretation 1
Automatic semantic interpretation 2
Semantic MEDLINE 1
Semantic MEDLINE 2
SemRep: Extract predication 1
SemRep: Extract predication 2
SemRep: Extract all predications
Abstraction summarization
Abstract

STUDY OBJECTIVES:

DESIGN:

MEASUREMENTS AND RESULTS:

A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men.
CONCLUSIONS:
Semantic Medline: Multi-Document Summarization and Visualization

Title Page
Information management application
Managing retrieval results 1
Managing retrieval results 2
Seamless integration of technologies
Semantic Medline Overview
Document selection
MEDLINE citations
Semantic interpretation 1
Semantic interpretation 2
Semantic predications
Summarization
Abstraction summarization
Salient semantic predications
Visualization
Informative graph
Semantic Medline Live
Home page
Search MEDLINE
Summarization options
Summarization results
More detail
Substance interactions
Acknowledgments

Semantic MEDLINE: An advanced information management application for biomedicine

Abstract
1 Introduction
   Figure 1 Semantic MEDLINE processing
2 Semantic interpretation of biomedical text: SemRep
   Figure 2 Example predication extracted from text by SemRep
   Figure 3 SemRep predications extracted from text on breast cancer. Predications outside the box are eliminated by automatic summarization
3. Semantic MEDLINE application
   Figure 4 Semantic MEDLINE graph
4. Conclusion
   Acknowledgement
"Both language and human thought are large, for feasibility we need to scale down the complexity of the process of semantic interpretation." Thomas C. Rindflesch, Ph.D., Lister Hill National Center for Biomedical Communication

MEDLINE (Medical Literature Analysis and Retrieval System Online) is a bibliographic database of life sciences and biomedical information. Compiled by the United States National Library of Medicine (NLM), MEDLINE is freely available on the Internet and searchable via PubMed and NLM's National Center for Biotechnology Information's Entrez system. The database contains more than 21.6 million records from 5,582 selected publications covering biomedicine and health from 1950 to the present. MEDLINE uses Medical Subject Headings (MeSH) for information retrieval.
MEDLINE functions as an important resource for biomedical researchers and journal clubs from all over the world. MEDLINE influences researchers in their choice of journals in which to publish: few biomedical researchers today consider publishing in a journal not indexed by MEDLINE, because other researchers would not find (or cite) their work.

Searching MEDLINE effectively is a learned skill; untrained users are sometimes frustrated with the large numbers of articles returned by simple searches. Counterintuitively, a search that returns thousands of articles is not guaranteed to be comprehensive. Unlike using a typical internet search engine, PubMed searching of MEDLINE requires a little investment of time. Using the MeSH database to define the subject of interest is one of the most useful ways to improve the quality of a search. Finding one article on the subject and clicking on the "Related Articles" link to get a collection of similarly classified articles can expand a search that yields few results.

PubMed is a free database accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics. The United States National Library of Medicine (NLM) at the National Institutes of Health maintains the database as part of the Entrez information retrieval system.

Semantic MEDLINE: the next PubMed? The NLM is working on a new interface for PubMed, including latest advances of semantic search (NLP and connected graphs).

“Semantic MEDLINE is a prototype Web application that summarizes MEDLINE citations returned by a PubMed search. Natural language processing is used to analyze salient content in titles and abstracts. This information is then presented in a graph that has links to the MEDLINE text processed.

Currently, the results from 35 PubMed searches (including a variety of disorders and drugs) are available to be processed. The 500 most recent citations (from the date of the search) are available for further processing by Semantic MEDLINE”.


Begin at the Search tab by selecting a search; then move to the Summarize tab. Choose a summary type to specify the point of view of the summary (Treatment of Disease, Substance Interactions, Diagnosis, or Pharmacogenomics). After selecting the topic of the summary, click the Summarize and Visualize button. The graph appears below. Right click on an edge to display a MEDLINE citation.

Tom Rndflesch and Fredrik Salvesen gave a presentation at the recent Government BIG DATA Symposium, June 16-17, 2014. See slides below.

MORE TO BE ADDED
Effectively Exploiting Big Data with Semantics: A Pilot Project

Thomas C. Rindflesch, National Library of Medicine
Fredrik Salvesen, YarcData
Slide 3 Disclaimer

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.

Slide 4 Background 1

Background

- Basic biomedical research is crucial to medicine
  - Complexity of molecular pathophysiology
  - Challenges development of new therapies
- Big data can facilitate progress
  - MEDLINE: Biomedical research literature
- Need effective, automatic access to information in this data source
Slide 5 Background 2

Background

• Basic biomedical research is crucial to medicine
  – Complexity of molecular pathophysiology
  – Challenges development of new therapies
• Big data can facilitate progress
  – MEDLINE: Biomedical research literature
• Need effective, automatic access to information in this data source
  – Semantic processing

Slide 6 Challenges

Challenges

• All data must be searched simultaneously to discover hidden relationships
• Response time needs to be improved over that available with commodity hardware
• Need a solution that can support 50 billion triples
  – Electronic medical record
  – Structured biomedical data
  – Web content
Slide 7 Solution

Solution

- Convert data to RDF graph database
- Use purpose build hardware
  - YarcData’s Urika graph appliance
  - Designed for graph database discovery
- Support large expansion, real time response, and limited performance degradation
- Pilot project for proof of concept

Slide 8 Pilot project components

Pilot project components

- Data and metadata in RDF
  - Nearly 24 million MEDLINE citations
  - Nearly 70 million semantic predicates
  - Both converted to 2.2 billion RDF triples
- Effective computing infrastructure
  - YarcData Urika graph appliance
- Application to manipulate semantic content of text
  - Semantic MEDLINE
Slide 9 The YarcData approach

The YarcData approach

- Research challenge:
  - uRIKA
  - Large Shared Memory Architecture
    - Up to 122 TB
  - Transportable Multi-Core
    - Up to 128 Threads
  - Scalable I/O
    - Up to 550 MB/s per hour

Real-time, interactive analytics on large graph problems

Slide 10 Semantic processing: SemRep

Semantic processing: SemRep

- Developed at National Library of Medicine
- Depends on domain knowledge
  - Unified Medical Language System (UMLS)
- Computable representation of meaning
  - Semantic predications
Slide 11 SemRep: semantic predication

SemRep: semantic predication

Exemestane after non-steroidal aromatase inhibitors for postmenopausal women with advanced breast cancer

Aromatase Inhibitor  TREATS  Breast Carcinoma

Metathesaurus Concept  Semantic Network Relation  Metathesaurus Concept

Unified Medical Language System

Slide 12 Web application: Semantic MEDLINE

Web application: Semantic MEDLINE

- Uses nearly 70 million semantic predications
  - From all of MEDLINE
- To guide the user through content
- Exploits existing IR system
  - PubMed
- Displays results as an interactive graph
Slide 13 Semantic MEDLINE overview

Semantic MEDLINE overview

- Document retrieval
- MEDLINE citations
- SemRep: semantic processing
- Semantic predications
- Automatic summarization
- Graphical summary
- Biomedical information management

Slide 14 Use case: Inflammation and cancer

Use case: Inflammation and cancer

- With some exceptions, cancer therapy is not effective
- Scientific basis
  - Traditionally: kill cancer cells
  - More recently: manipulate non-cancer cells (immune system)
- Goal: look for trends in cancer immunotherapy
Slide 15 SemMed: semantic predications

Slide 16 Semantic predications as a graph
Slide 17 Cytokine CAUSES Tumorigenesis

Slide 18 MEDLINE citation
Slide 19 Exploiting semantic processing

Exploiting semantic processing

- Research
  - Literature-based discovery (LBD)
    - Hypothesis generation
  - Discovery browsing
    - Investigate salient aspects of a topic
- Trends
  - Discern trends: Where is research headed?
  - Guide trends: Where should it be headed?

Slide 20 Acknowledgments

Acknowledgments

- Michael J. Cairelli, D.O.
- Marcelo Fiszman, M.D., Ph.D.
- Halil Kilicoglu, Ph.D.
- Graciela Roseblat, Ph.D.
- Dongwook Shin, Ph.D.
- YarcData team
  - Aaron Bossert
  - Ted Slater
  - Tim White
  - Fredrik Salvesen

Slide 21 Additional information

Slides, including 7 minute live demo: http://www.youtube.com/watch?v=ShfI4SNzNO4

Slides, including 20 minute live demo: http://www.youtube.com/watch?v=6frNAmPD0mo

Thomas C. Rindflesch (tcr@nlm.nih.gov) and Fredrik Salvesen (fredrik@salvesen.me)
Spotfire Dashboard

For Internet Explorer Users and Those Wanting Full Screen Display Use: Web Player Get Spotfire for iPad App

Media, iframe, embed and object tags are not supported inside of a PDF.

Spotfire Network Analytics

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
Semantic MedLine Prototype

Home


Note: If you wish to use the current system with your own searches, please contact Dr. Thomas Rindflesch trindflesch@mail.nih.gov

Semantic MEDLINE is a prototype Web application that summarizes MEDLINE citations returned by a PubMed search. Natural language processing is used to analyze salient content in titles and abstracts. This information is then presented in a graph that has links to the MEDLINE text processed.
Currently, the results from 35 PubMed searches (including a variety of disorders and drugs) are available to be processed. The **500 most recent citations** (from the date of the search) are available for further processing by Semantic MEDLINE.

Begin at the Search tab by selecting a search; then move to the Summarize tab. Choose a summary type to specify the point of view of the summary (Treatment of Disease, Substance Interactions, Diagnosis, or Pharmacogenomics). After selecting the topic of the summary, click the Summarize and Visualize button. The graph appears below. Right click on an edge to display a MEDLINE citation.

---

**Help**


**SEARCH PAGE**

After you select the MEDLINE search you wish to run, click the Search button. Clicking the Reset button erases all previously-set search options.

The MEDLINE search results are displayed at the bottom of the Search page. MEDLINE results appear as a brief display of the PMID (PubMed Unique Identifier) and citation title. Clicking on the PMID or the title takes you directly to an AbstractPlus display of the citation in PubMed. Clicking on the PMID column header sorts the PMIDs in ascending or descending numeric order.

Search results are viewable in groups of 20. You can navigate through the search results by clicking the [First/Prev] group of 20 numbers [1,2,3,4,5][Next/Last] options.

Click the Summarize tab to begin linguistic processing of the selected MEDLINE search results.

**SUMMARIZATION PAGE**

The top of the Summarize page shows a brief description of search parameters from the current session, including Search Term, Source, Dates, and the number of predications extracted from the MEDLINE citations.

A predication is a formal representation of a relationship asserted in text. A predication is a triple consisting of two arguments joined by a relation. For example, the predication "Aspirin-TREATS-Headache" has arguments "Aspirin" and "Headache," and relation "TREATS."

Summarization is a linguistic process that extracts key information (predications) from the text (title and abstract) of MEDLINE citations. Extracted information centers around a main topic and is represented as predications visualized as a graph.

You can specify Summarization Options for your particular interests. Summary Type specifies the point of view of the summarization process and determines the kind of information contained in the summary. Summary Types available are: Treatment of Disease, Substance Interactions, Diagnosis, and Pharmacogenomics. The default is Treatment of Disease.
Treatment of Disease concentrates on the use of therapeutic interventions (drugs and procedures) for disorders. Substance Interactions refers to how substances (including drugs) affect other substances and relevant diseases. Diagnosis highlights procedures for identifying diseases by associated signs and symptoms. Pharmacogenomics refers to how genes affect a person's response to drugs. Pharmacogenomics combines pharmacology (the science of drugs) and genomics (the study of genes and their functions) to develop effective, safe medications and doses tailored to a person's genetic makeup.

The second Summarization Option is a checkbox to "Keep only the most frequent predications." You can leave the checkbox on (default) or turn it off by uncircling the green checkmark. This will omit less frequently occurring predications from the summary. Keeping only the most frequent predications (with the box checked) helps focus the summary.

The final Summarization Option is to select the central concept of the summary. This is chosen from a list of arguments from all predications in the summary. This list appears in descending order of frequency of occurrence and changes depending on the Summary Type specified.

Click the Summarize and Visualize button to display the graph representing the summarized information.

MULTI-PANEL WINDOW

The graph appears as one of the panels of a multi-panel window at the bottom of the Summarize page. A summary statement indicating how many predications (including unique predications) were extracted from the MEDLINE citations processed appears above the window. The window provides information about the graph as well as search and navigation facilities.

Graph Info

Click the Graph Info box to display statistics about the graph. A small pop-up box appears showing counts for Concept (argument), Predication (relation), Visible Concept, and Visible Predication. Invisible arguments and relations result from hiding predications in the graph. (See Relations Panel below.)

Search Concept

Type in an argument (or substring) and press enter to locate an argument in the graph. If found, the argument will be highlighted in red.

Zoom

Zoom is a horizontal slidebar that spreads out the graph so you have more room to navigate. Move the slide bar to the right to spread out; move to the left to collapse.

Stop

Click the Stop button to freeze graph. This is useful when large graphs are displayed.
PANELS

The 4 panels of the window are: the Graph (main panel), Citation (below the Graph), the Relations (upper right), and Information (lower right).

Graph Panel

The graph contains the predications of the summary. Nodes represent arguments, and arcs represent relations. The central concept of the summary appears as a node with text in red. Nodes are colored-coded into 15 high-level groups representing biomedical categories such as Chemicals and Drugs and Disorders (Figure 1).

Figure 1. High-Level Groups for Concepts

Right click on a node to display active links to other National Library of Medicine resources, such as the Unified Medical Language System Metathesaurus, Entrez Gene, OMIM (Online Mendelian Inheritance in Man), and GHR (Genetics Home Reference). To use these links, allow pop-ups in your Web browser. Links are unavailable if grayed out. Clicking on a node also causes information about it to appear in the Information Panel (see below).

The arcs representing relations between arguments are color-coded to indicate the name of the relation (see Relations Panel below). The direction towards which the arc points indicates the direction of the relation between the two arguments. Clicking on an arc causes it to change color to black indicating that you are navigating from one predication to another.

Right-click on an arc and the Display Citations button appears. Left-click on the Display Citations button and the Citation panel (below Graph panel) shows the sentence from the MEDLINE citation that generated the predication.
Citation Panel

The Citation panel contains the sentence in a MEDLINE citation from which a selected relation was extracted. The sentence is highlighted in yellow. Information about the citation includes PMID, date of publication (DP), Title (TI), and Abstract (AB). The PMID is hyperlinked to the citation in PubMed.

Relations Panel

The Relations panel provides a legend for the colors of the arcs representing names of relations in the graph. (See Relations Definitions below.) You can toggle relations on and off in the graph by clicking the check box in the Relations panel. If you toggle a relation off, the corresponding arc in the graph disappears along with associated nodes. A little red box in the upper right hand corner of a node indicates the number of hidden nodes.

Information Panel:

The Information panel (lower right panel) changes as you click between nodes (arguments) and arcs (relations) to display linguistic and statistical information about the active argument or predication.

RELATION DEFINITIONS

ADMINISTERED_TO

To be given to an entity, when no assertion is made that the substance is being given as treatment.

AFFECTS

Produces a direct effect on. Implied here is the altering or influencing of an existing condition, state, situation, or entity. This includes has a role in, alters, influences, predisposes, catalyzes, stimulates, regulates, depresses, impedes, enhances, contributes to, leads to, and modifies.

ASSOCIATED_WITH

Has a significant or salient relationship to.

AUGMENTS

To expand or stimulate a process.

CAUSES

Brings about a condition or an effect. Implied here is that an agent, such as for example, a pharmacologic substance or an organism, has brought about the effect. This includes induces, effects, evokes, and etiology.
COEXISTS_WITH

Occurs at the same time as, together with, or jointly. This includes is co-incident with, is concurrent with, is contemporaneous with, accompanies, coexists with, and is concomitant with.

CONVERTS_TO

To change in form from arg1 to arg2, (both substances)

COMPLICATES

Causes to become more severe or complex or results in adverse effects.

DIAGNOSES

Distinguishes or identifies the nature or characteristics of.

DISRUPTS

Alters or influences an already existing condition, state, or situation. Produces a negative effect on.

INHIBITS

Decreases, limits, or blocks the action or function of something.

INTERACTS_WITH

Acts, functions, or operates together with.

ISA

The basic hierarchical link in the UMLS Semantic Network. If one item "isa" another item then the first item is more specific in meaning than the second item.

LOCATION_OF

The position, site, or region of an entity or the site of a process.

MANIFESTATION_OF

That part of a phenomenon which is directly observable or concretely or visibly expressed, or which gives evidence to the underlying process. This includes expression of, display of, and exhibition of.
PART_OF
Composes, with one or more other physical units, some larger whole. This includes component of, division of, portion of, fragment of, section of, and layer of.

PREDISPOSES
To be involved or be a risk to a disorder, pathology, or condition. The agent such as substance or finding is not explicitly involved in the etiology of the disorder, pathology, or condition.

PREVENTS
Stops, hinders or eliminates an action or condition.

PROCESS_OF
Action, function, or state of.

STIMULATES
Increases or facilitates the action or function of something

TREATS
Applies a remedy with the object of effecting a cure or managing a condition.

Semantic Medline

Account Information
Here is Semantic Medline URL and your account info.
Account: bniemann
Password:

SemMed Documentation (PDF)

Home
Search Options:

Most Recent: 100  Start Date: 01/01/1900
End Date: 01/31/2012
PubMed Limits: Show

Search


My Note: Need to Click on Search Button and See MedLine Citations below that have links to full page Citation and popup Abstract.

Query: Alzheimer

Search Options:

Most Recent: 100  Start Date: 01/01/1900
End Date: 01/31/2012
PubMed Limits: Hide

- Research Support, U.S. Gov't, Non-P.H.S.
- Research Support, U.S. Gov't, P.H.S.
- Retraction of Publication
- Scientific Integrity Review
- Technical Report
- Twin Study
- Validation Studies

Ages
- All Infant: birth-23 months
- All Child: 0-18 years
- All Adult: 19+ years
- Newborn: birth-1 month
Infant: 1-23 months
Preschool Child: 2-5 years
Child: 6-12 years
Adolescent: 13-18 years
Adult: 19-44 years
Middle Aged: 45-64 years
Middle Aged + Aged: 45+ years
Aged: 65+ years
80 and over: 80+ years

Found 500 citations. Showing 1 to 20
Pages: Prev 1 | 2 | 3 | ... | 25 | Next

Citations

Lott I.
Antioxidants in Down syndrome.
22006998

Pocernich C, Butterfield D.
Elevation of glutathione as a therapeutic strategy in Alzheimer disease.
22015471

Abnormal temporal dynamics of visual attention in Alzheimer's disease and in dementia with Lewy bodies.
22130206

Microbleeds relate to altered amyloid-beta metabolism in Alzheimer's disease.
22118945

de Waal H, Stam C, de Haan W, van Straaten E, Scheltens P, van der Flier W.
Young Alzheimer patients show distinct regional changes of oscillatory brain dynamics.
22118944

MTHFR (677 and 1298) and IL-6-174 G/C genes in pathogenesis of Alzheimer's and vascular dementia and their epistatic interaction.
22015309

Pietropaolo S, Delage P, Lebreton F, Crusio W, Cho Y.
Early development of social deficits in APP and APP-PS1 mice.
22014620

Serum paraoxonase activity is associated with variants in the PON gene cluster and risk of Alzheimer disease.
20980077
Li D, Zhao H, Kranzler H, Oslin D, Anton R, Farrer L, Gelernter J.
Association of COL25A1 with Comorbid Antisocial Personality Disorder and Substance Dependence. 22297151

Novel S-acyl glutathione derivatives prevent amyloid oxidative stress and cholinergic dysfunction in Alzheimer disease models. 22326489

RNA modifications by oxidation: A novel disease mechanism? 22306201

Dementia is associated with Insulin Resistance in patients with Parkinson's Disease. 22265943

Kaden D, Munter L, Reif B, Multhaup G.
The amyloid precursor protein and its homologues: Structural and functional aspects of native and pathogenic oligomerization. 21459473

Aydin D, Weyer S, Müller U.
Functions of the APP gene family in the nervous system: insights from mouse models. 21931985

Human neural stem cells over-expressing choline acetyltransferase restore cognition in rat model of cognitive dysfunction. 22245157

Ideno Y, Takayama M, Hayashi K, Takagi H, Sugai Y.
Evaluation of a Japanese version of the Mini-Mental State Examination in elderly persons. 22122408

Leoutsakos J, Muthen B, Breitner J, Lyketsos C.
Effects of non-steroidal anti-inflammatory drug treatments on cognitive decline vary by phase of pre-clinical Alzheimer disease: findings from the randomized controlled Alzheimer's Disease Anti-inflammatory Prevention Trial. 21560159

Daliri M.
Automated Diagnosis of Alzheimer Disease using the Scale-Invariant Feature Transforms in Magnetic Resonance Images. 21584770
Jellinger K.
CSF biomarkers in different phenotypes of Parkinson disease.
22065209

Short latency afferent inhibition differs among the subtypes of mild cognitive impairment.
22016008

Summarization


My Note: Get the Predicate Data Set as a Table!

(Query: Alzheimer, Source: Medline, Most Recent: 500, Start Date: 01/01/1900, End Date: 01/31/2012, 500 citations retrieved., 2616 predications extracted.)

Options:

Summary Type: Treatment of Disease

☐ More Relations
Select a UMLS concept to summarize on:
<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's Disease</td>
<td>647</td>
</tr>
<tr>
<td>Dementia</td>
<td>97</td>
</tr>
<tr>
<td>Neurodegenerative Disorders</td>
<td>51</td>
</tr>
<tr>
<td>Parkinson Disease</td>
<td>38</td>
</tr>
<tr>
<td>Atrophic</td>
<td>31</td>
</tr>
<tr>
<td>Oxidative Stress</td>
<td>29</td>
</tr>
<tr>
<td>Down Syndrome</td>
<td>20</td>
</tr>
<tr>
<td>Senile Plaques</td>
<td>18</td>
</tr>
<tr>
<td>Nerve Degeneration</td>
<td>17</td>
</tr>
<tr>
<td>Toxic effect</td>
<td>17</td>
</tr>
<tr>
<td>Dementia, Vascular</td>
<td>16</td>
</tr>
<tr>
<td>Amyloid deposition</td>
<td>14</td>
</tr>
<tr>
<td>Huntington Disease</td>
<td>14</td>
</tr>
<tr>
<td>Alzheimer Disease, Late Onset</td>
<td>11</td>
</tr>
<tr>
<td>Cerebral atrophy</td>
<td>10</td>
</tr>
<tr>
<td>Cerebrovascular accident</td>
<td>10</td>
</tr>
<tr>
<td>Demented</td>
<td>10</td>
</tr>
<tr>
<td>Familial Alzheimer's disease</td>
<td>10</td>
</tr>
<tr>
<td>Creutzfeldt-Jakob Syndrome</td>
<td>9</td>
</tr>
<tr>
<td>Inflammation</td>
<td>9</td>
</tr>
<tr>
<td>Memory impairment</td>
<td>9</td>
</tr>
<tr>
<td>Depressive disorder</td>
<td>8</td>
</tr>
<tr>
<td>Diabetes</td>
<td>8</td>
</tr>
<tr>
<td>Neurofibrillary Tangles</td>
<td>8</td>
</tr>
<tr>
<td>Pathologic Processes</td>
<td>8</td>
</tr>
<tr>
<td>Degenerative abnormality</td>
<td>7</td>
</tr>
<tr>
<td>Infarction</td>
<td>7</td>
</tr>
<tr>
<td>Alzheimer Disease, Early Onset</td>
<td>6</td>
</tr>
<tr>
<td>Cerebral hemisphere hemorrhage</td>
<td>6</td>
</tr>
<tr>
<td>Disease Progression</td>
<td>6</td>
</tr>
<tr>
<td>PMID</td>
<td>Sentence</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>20359781</td>
<td>In addition to atrophy of mesial temporal lobe structures critical for memory function, white matter projections to the hippocampus may be compromised in individuals with mild Alzheimer's disease (AD), thereby compounding the memory difficulty.</td>
</tr>
<tr>
<td>21351296</td>
<td>Although depressive symptoms may hasten conversion from mild cognitive impairment (MCI) to dementia, depression-related conversion is not likely to be mediated by evolution of the AD pathological process.</td>
</tr>
<tr>
<td>21384431</td>
<td>The subsetting analysis of the ADAS-cog combined database aimed at selecting the scale items showing no worsening at study end compared to baseline due to memantine treatment in mild AD (Mini-Mental State Examination (MMSE &gt;19)) patients.</td>
</tr>
<tr>
<td>21384431</td>
<td>The approach was applied in a post-hoc analysis of ADAS-cog results from two randomized, placebo-controlled and double-blind clinical trials with memantine in mild to moderate Alzheimer's disease (AD).</td>
</tr>
<tr>
<td>21502851</td>
<td>Little is known regarding factors associated with soluble amyloid beta peptide (Abeta) concentrations in humans at late midlife, when Abeta is likely most critical to Alzheimer disease pathogenesis.</td>
</tr>
</tbody>
</table>
Current diagnostic tools can detect neuronal degeneration in AD only after irreversible damage that already existed.

METHODS: 113 participants with probable AD were assessed for severity of disease, cognitive and functional impairment.

Alzheimer's disease (AD) is the most common type of dementia.

In a separate autopsy sample of individuals with AD (N=17) and healthy controls (N=4), we examined the association between antemortem clusterin concentration in plasma and postmortem levels in the superior temporal gyrus, hippocampus and cerebellum.

PARTICIPANTS: A total of 401 elderly participants in the Alzheimer's Disease Neuroimaging Initiative who were cognitively normal, who had mild cognitive impairment, or who had AD dementia.

The proportion of participants with an abnormal total tau level increased from baseline to 12 months in cognitively normal participants (P = .05) but not in participants with mild cognitive impairment or AD dementia.

Alzheimer's disease (AD) is a common neurodegenerative disorder of late life with a complex genetic basis.
characterized by progressive memory loss and cognitive disturbance.

PARTICIPANTS: Individuals with probable AD and detailed parental history (n = 5370).

BACKGROUND: Oxidative stress (OS) may be involved in the neurodegenerative process in Alzheimer's disease (AD).

Alzheimer Disease (AD) is the most prevalent form of dementia and the sixth leading cause of death in developed world.

Alzheimer's disease is the most devastating neurodegenerative disorder in the elderly, yet treatment options are severely limited.

Alzheimer disease (AD) is a complex neurodegenerative disease.

Semantic MEDLINE Documentation

Source: Semantic MedLine Documentation (PDF)

Introduction

Semantic MEDLINE is a prototype Web application that integrates PubMed searching, advanced natural language processing, automatic summarization, and visualization into a single Web portal. The application is intended to help manage the results of PubMed searches by identifying semantic predications in the citations retrieved. These predications constitute computable knowledge accessible to further manipulation, including condensation by automatic summarization. Summarized predications are visualized as an informative graph with links to the original MEDLINE citations. Convenient access is also provided to additional relevant knowledge resources, including Entrez Gene, the Genetics Home Reference, and the Unified Medical Language System Metathesaurus.

Semantic MEDLINE is organized around two tabs: Search and Summarization. Begin at the Search tab by issuing a PubMed query; then move to the Summarization tab. Choose a summary type to specify content focus of the summary (Treatment of Disease, Substance Interactions, Diagnosis, or Pharmacogenomics). After selecting a topic, click the Summarize and then the Visualize button to display a graph. Click on an arc in the graph to display MEDLINE citation text.

Search Tab

You can issue any PubMed search in the Query box. Any date range can be entered between 01/01/1999 (in the Start Date box) and the end date (in the End Date box). The number of citations to be retrieved is specified in the Most Recent drop-down menu. The default is 500. PubMed Limits provides access to most of the PubMed options that are
available to focus your search. After you enter the query (and specify search options), click the Search button. Clicking the Reset button erases all previously-set search options.

Search results appear at the bottom of the page as a brief display of the authors, title, and PMID (PubMed Unique Identifier.) Clicking on the list of authors takes you directly to an AbstractPlus display of the citation in PubMed. Clicking on Abstract (on the right) will display the text of the abstract. Search results are viewable in groups of 20. You can navigate through the search results by clicking the [Prev] group of 20 numbers [1 | 2, | 3 | 4 | 5] Next] options.

**Summarization Tab**

The top of the Summarization page shows a brief description of search parameters from the current search, including query, number of citations retrieved, and number of predications extracted.

A predication is a formal representation of a relationship asserted in text and consists of two arguments (subject and object) joined by a relation (or predicate). For example, the predication “Tamoxifen TREATS Breast Carcinoma” has subject “Tamoxifen,” object “Breast Carcinoma,” and relation “TREATS.” This predication could have been retrieved from a sentence such as “In this study, tamoxifen was used in the treatment of breast carcinoma” found in a MEDLINE abstract. Predications extracted from text serve as a representation of the content of that text and form the basis for further processing.

Summarization identifies key predications from the text being processed. These predications center around a main topic and are visualized as a graph. You select summarization options for your particular interests. Summary Type specifies the content focus of the summarization process and determines the kind of information contained in the summary. Summary Types available are: Treatment of Disease, Substance Interactions, Diagnosis, and Pharmacogenomics. The default is Treatment of Disease, which concentrates on the use of therapeutic interventions (drugs and procedures) for disorders. Substance Interactions refers to how substances (including drugs) affect other substances and relevant diseases. Diagnosis highlights procedures for identifying diseases by associated signs and symptoms. Pharmacogenomics refers to how genes affect a person’s response to drugs.

The second summarization option is a checkbox to include or not include More Relations. Unchecked, the option omits less frequently occurring predications from the summary.

The final summarization option is to select a UMLS concept as the topic of the summary. This is chosen from a list of concepts from all predications extracted from the text being processed. The list appears in descending order of frequency of occurrence. The semantic class of available concepts changes depending on the Summary Type specified:

- Treatment of Disease: Disorders
- Substance Interactions: Substances
- Diagnosis: Disorders
- Pharmacogenomics: Disorders and Substances

Click the Summarize button to display the predications representing the summarized information. Predications are displayed at the bottom of the page as a list that includes the PMID and the sentence from which the predication was extracted, followed by the Subject, Predicate, and Object that constitute the predication. Clicking the PMID in this list takes you to the text of the MEDLINE citation (title or abstract). Predications are viewable in groups of 20 by clicking on the navigation aids under the message reporting on the number of predications found. Individual predications can be inspected with this facility, but an overview of the summarized predications can be seen in the graph generated by clicking the Visualize button that appears under the Summarize button.

**Visualization Page**

Clicking the Visualize button displays the visualization page, which includes a graph representing the summarized information contained in the predications. (Install the Adobe Flash plug-in from [http://www.adobe.com/products/](http://www.adobe.com/products/).)
Graph Control Pane

The overall shape of the graph can be changed using the Layout pull-down menu. There are four choices available: Spring, Node-Link Tree, Circle, and Radial. The default is Node-Link Tree.

- Spring: The nodes push on each other until they stabilize around the topic of the summary. (This format should be avoided for large graphs.)
- Node-Link Tree: The nodes and arcs extend in a tree to the right. The summary topic is on the left.
- Circle: The nodes are arranged in a circle.
- Radial: Similar to Spring, although more stable. The topic of the summary is centrally located.

Magnifying Glass

Use the magnifying glass button to increase or decrease the size of the graph. You can also use the wheel on your mouse to change the graph size. (Note: If the graph seems to disappear, reduce the size of the graph until it reappears, and then move it to the center of the page.)

Stop/Restart

Stop/Restart toggles the movement of the nodes when the Spring graph layout is displayed. Stop/Restart is useful when large graphs are displayed.

Graph Pane

The graph represents the summarized predications from the titles and abstracts of the citations retrieved. Nodes represent arguments (subject and object), and arcs represent relations (or predicates). The direction towards which the arc points indicates the direction of the relation between the two arguments. Both nodes and arcs are color coded to represent semantic information inherent in the predication. Nodes are colored coded into 15 high-level semantic groups representing the biomedical categories of the arguments. These include, for example, Chemicals and Drugs and Disorders (Figure 1).

Figure 1. Colors of High Level Groups for Concepts

![Colors of High Level Groups for Concepts](http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline)

The arcs representing relations between arguments are color-coded to indicate the name of the relation (see Relation Labels pane below). For example, the relation “Tamoxifen TREATS

Breast Carcinoma” will have a light yellow node for Tamoxifen (Chemicals & Drugs) and a light pink node for Breast Carcinoma (Disorders). A blue arc representing the relation TREATS will point from Tamoxifen to Breast Carcinoma.
Clicking on a node (when the cursor appears as a hand) displays additional information about the argument in the upper part of the Information pane (see Concept Information below). Clicking on an arc (when the cursor appears as a hand) provides more information about the predication, including the text from which it was extracted. The graph can be moved by clicking on white space outside a node or arc. Individual nodes can be moved by clicking on them; this is useful if the names of the nodes overlap. The layout of the graph can be changed by making a choice in the Graph Control pane at the bottom of the page.

Information Pane

Concept Information

Clicking on a node displays information about the node (argument) in the upper right of the Information Pane. The Concept Information comes from the UMLS (Unified Medical Language System, http://www.nlm.nih.gov/research/umls...mentation.html): CUI (Concept Unique Identifier) and Semantic Type (UMLS category the concept belongs to). See below for definitions of UMLS semantic types. The number of times the concept appears in the graph is also given. Buttons under Concept Information display active links to other National Library of Medicine resources, such as the UMLS Metathesaurus, GHR (Genetics Home Reference), OMIM (Online Mendelian Inheritance in Man), and Entrez Gene. To use these links, allow pop-ups in your Web browser. Links are unavailable if grayed out.

Relationship Information

Clicking on an arc in the graph causes information about the relation to appear in the Information pane (lower right.) The Relationship Information displays the subject, relation (predicate) and object of the predication selected as well as number of occurrences. “No. Predications:” gives the overall frequency of the predication in the text being processed, while “No. Citations:” indicates the number of citations in which the predication appears. A selected predication may appear in more than one citation.

Clicking the Citations button opens a window containing the text from which the predications were extracted. The PMID, title, and abstract are displayed. The PubMed ID is linked to PubMed. The specific sentence from which the predication was extracted is highlighted in orange.

Relation Labels

Clicking on Relation Labels at the bottom of the Information pane provides a legend for the colors of the arcs representing relations in the graph (see Relation Definitions below). You can toggle relations on and off in the graph by clicking the check box in the Relations section of the Relation Labels pane. If you toggle a relation off, the corresponding arcs in the graph disappear along with associated nodes.

Search

Clicking Search (bottom right) allows you to locate a specific concept in the graph. Type in a term (or substring) and click the Search button. Concepts matched with the search terms are listed in the box below the search box. When a matched term is selected, the graph repositions so that the selected term is central.

SEMANTIC RELATION (PREDICATE) DEFINITIONS
ADMINISTERED_TO
Given to patient, when no assertion is made that the substance is being given as treatment.

AFFECTS
Produces a direct effect on. Implied here is the altering or influencing of an existing condition, state, situation, or entity. This includes has a role in, alters, influences, predisposes, catalyzes, stimulates, regulates, depresses, impedes, enhances, contributes to, leads to, and modifies.

ASSOCIATED_WITH
Has a relationship to (gene-disease relation).

AUGMENTS
Expands or stimulates a process.

CAUSES
Brings about a condition or an effect. Implied here is that an agent, such as for example, a pharmacologic substance or an organism, has brought about the effect. This includes induce, affects, evokes, and etiology.

COEXISTS_WITH
Occurs together with, or jointly.

CONVERTS_TO
Changes from one form to another (both substances).

COMPLICATES
Causes to become more severe or complex, or results in adverse effects.

DIAGNOSES
Distinguishes or identifies the nature or characteristics of.

DISRUPTS
Alters or influences an already existing condition, state, or situation. Produces a negative effect on.

INHIBITS
Decreases, limits, or blocks the action or function of (substance interaction).

INTERACTS_WITH
Substance interaction.
ISA
The basic hierarchical link in the UMLS Semantic Network. If one item "is a" another item, then the first item is more specific in meaning than the second item.

LOCATION_OF
The position, site, or region of an entity or the site of a process.

MANIFESTATION_OF
That part of a phenomenon which is directly observable or concretely or visibly expressed, or which gives evidence to the underlying process. This includes expression of, display of, and exhibition of.

METHOD_OF
The manner and sequence of events in performing an act or procedure.

OCCURS_IN
Has incidence in a group or population.

PART_OF
Composes, with one or more other physical units, some larger whole. This includes component of, division of, portion of, fragment of, section of, and layer of.

PRECEDES
Occurs earlier in time. This includes antedates, comes before, is in advance of, predates, and is prior to.

PREDISPOSES
To be a risk to a disorder, pathology, or condition.

PREVENTS
 Stops, hinders or eliminates an action or condition.

PROCESS_OF
Disorder occurs in (higher) organism.

PRODUCES
Brings forth, generates or creates. This includes yields, secretes, emits, biosynthesizes, generates, releases, discharges, and creates.
STIMULATES
Increases or facilitates the action or function of (substance interaction).

TREATS
Applies a remedy with the object of effecting a cure or managing a condition.

USES
Employs in the carrying out of some activity. This includes applies, utilizes, employs, and avails.

SEMANTIC TYPE DEFINITIONS (organized by node semantic groups)

Activities and Behaviors

Activity
An operation or series of operations that an organism or machine carries out or participates in.

Behavior
Any of the psycho-social activities of humans or animals that can be observed directly by others or can be made systematically observable by the use of special strategies.

Daily or Recreational Activity
An activity carried out for recreation or exercise, or as part of daily life.

Event
A broad type for grouping activities, processes and states.

Governmental or Regulatory Activity
An activity carried out by officially constituted governments, or an activity related to the creation or enforcement of the rules or regulations governing some field of endeavor.

Individual Behavior
Behavior exhibited by a human or an animal that is not a direct result of interaction with other members of the species, but which may have an effect on others.

Machine Activity
An activity carried out primarily or exclusively by machines.
Occupational Activity
An activity carried out as part of an occupation or job.

Social Behavior
Behavior that is a direct result or function of the interaction of humans or animals with their fellows. This includes behavior that may be considered anti-social.

Anatomy

Anatomical Structure
A normal or pathological part of the anatomy or structural organization of an organism.

Body Location or Region
An area, subdivision, or region of the body demarcated for the purpose of topographical description.

Body Part, Organ, or Organ Component
A collection of cells and tissues which are localized to a specific area or combine and carry out one or more specialized functions of an organism. This ranges from gross structures to small components of complex organs. These structures are relatively localized in comparison to tissues.

Body Space or Junction
An area enclosed or surrounded by body parts or organs or the place where two anatomical structures meet or connect.

Body Substance
Extracellular material, or mixtures of cells and extracellular material, produced, excreted, or accreted by the body. Included here are substances such as saliva, dental enamel, sweat, and gastric acid.

Body System
A complex of anatomical structures that performs a common function.

Cell
The fundamental structural and functional unit of living organisms.

Cell Component
A part of a cell or the intercellular matrix, generally visible by light microscopy.
Embryonic Structure
An anatomical structure that exists only before the organism is fully formed; in mammals, for example, a structure that exists only prior to the birth of the organism. This structure may be normal or abnormal.

Fully Formed Anatomical Structure
An anatomical structure in a fully formed organism; in mammals, for example, a structure in the body after the birth of the organism.

Tissue
An aggregation of similarly specialized cells and the associated intercellular substance. Tissues are relatively non-localized in comparison to body parts, organs or organ components.

Chemicals and Drugs

Amino Acid, Peptide, or Protein
Amino acids and chains of amino acids connected by peptide linkages.

Antibiotic
A pharmacologically active compound produced by growing microorganisms which kill or inhibit growth of other microorganisms.

Biologically Active Substance
A generally endogenous substance produced or required by an organism, of primary interest because of its role in the biologic functioning of the organism that produces it.

Biomedical or Dental Material
A substance used in biomedicine or dentistry predominantly for its physical, as opposed to chemical, properties. Included here are biocompatible materials, tissue adhesives, bone cements, resins, toothpastes, etc.

Carbohydrate
A generic term that includes monosaccharides, oligosaccharides, and polysaccharides as well as substances derived from monosaccharides by reduction of the carbonyl group (alditols), by oxidation of one or more terminal group to carboxylic acids, or by replacement of one or more hydroxy groups by a hydrogen atom, an amino group, a thiol group or similar heteroatomic groups. It also includes derivatives of these compounds. Included here are sugar phosphates. Excluded are glycolipids and glycoproteins.
Chemical
Compounds or substances of definite molecular composition. Chemicals are viewed from two distinct perspectives in the network, functionally and structurally. Almost every chemical concept is assigned at least two types, generally one from the structure hierarchy and at least one from the function hierarchy.

Chemical Viewed Functionally
A chemical viewed from the perspective of its functional characteristics or pharmacological activities.

Chemical Viewed Structurally
A chemical or chemicals viewed from the perspective of their structural characteristics. Included here are concepts which can mean either a salt, an ion, or a compound (e.g., "Bromates" and "Bromides").

Clinical Drug
A pharmaceutical preparation as produced by the manufacturer. The name usually includes the substance, its strength, and the form, but may include the substance and only one of the other two items.

Eicosanoid
An oxygenated metabolite from polyunsaturated 20 carbon fatty acids including lipoxygenase and cyclooxygenase products and their synthetic analogs. This includes the prostaglandins and thromboxanes.

Element, Ion, or Isotope
One of the 109 presently known fundamental substances that comprise all matter at and above the atomic level. This includes elemental metals, rare gases, and most abundant naturally occurring radioactive elements, as well as the ionic counterparts of elements (NA+, Cl-), and the less abundant isotopic forms. This does not include organic ions such as iodoacetate to which the type 'Organic Chemical' is assigned.

Enzyme
A complex chemical, usually a protein, that is produced by living cells and which catalyzes specific biochemical reactions. There are six main types of enzymes: oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases.

Hazardous or Poisonous Substance
A substance of concern because of its potentially hazardous or toxic effects. This would include most drugs of abuse, as well as agents that require special handling because of their toxicity.

Hormone
In animals, a chemical usually secreted by an endocrine gland whose products are released into the circulating fluid. Hormones act as chemical messengers and regulate various physiologic processes such as growth, reproduction, metabolism, etc. They usually fall into two broad classes, steroid hormones and peptide hormones.
Immunologic Factor
A biologically active substance whose activities affect or play a role in the functioning of the immune system.

Indicator, Reagent, or Diagnostic Aid
A substance primarily of interest for its use in laboratory or diagnostic tests and procedures to detect, measure, examine, or analyze other chemicals, processes, or conditions.

Inorganic Chemical
Chemical elements and their compounds, excluding the hydrocarbons and their derivatives (except carbides, carbonates, cyanides, cyanates and carbon disulfide). Generally inorganic compounds contain ionic bonds. Included here are inorganic acids and salts, alloys, alkalis, and minerals.

Lipid
An inclusive group of fat or fat-derived substances that are soluble in nonpolar solvents related to fatty acid esters, fatty alcohols, sterols, waxes, etc. Included in this group are the saponifiable lipids such as glycerides (fats and oils), essential (volatile) oils, and phospholipids.

Neuroreactive Substance or Biogenic Amine
An endogenous substance whose activities affect or play an important role in the functioning of the nervous system. Included here are catecholamines, neuroregulators, neurophysins, etc.

Nucleic Acid, Nucleoside, or Nucleotide
A complex compound of high molecular weight occurring in living cells. These are basically of two types, ribonucleic (RNA) and deoxyribonucleic (DNA) acids. Nucleic acids are made of nucleotides (nitrogen-containing base, a 5-carbon sugar, and one or more phosphate group) linked together by a phosphodiester bond between the 5’ and 3’ carbon atoms. Nucleosides are compounds composed of a purine or pyrimidine base (usually adenine, cytosine, guanine, thymine, uracil) linked to either a ribose or a deoxyribose sugar.

Organic Chemical
The general class of carbon-containing compounds, usually based on carbon chains or rings, and also containing hydrogen (hydrocarbons), with or without nitrogen, oxygen, or other elements in which the bonding between elements is generally covalent.

Organophosphorus Compound
An organic compound containing phosphorus as a constituent. Included here are organic phosphinic, phosphonic and phosphoric acid derivatives and their thiophosphorus counterparts. Excluded are phospholipids, sugar phosphates, phosphoproteins, nucleotides, and nucleic acids.

Pharmacologic Substance
A substance used in the treatment or prevention of pathologic disorders. This includes substances that occur naturally in the body and are administered therapeutically.
Receptor
A specific structure or site on the cell surface or within its cytoplasm that recognizes and binds with other specific molecules. These include the proteins on the surface of an immunocompetent cell that binds with antigens, or proteins found on the surface molecules that bind with hormones or neurotransmitters and react with other molecules that respond in a specific way.

Steroid
One of a group of polycyclic, 17-carbon-atom, fused-ring compounds occurring both in natural and synthetic forms. Included here are naturally occurring and synthetic steroids, bufanolides, cardanolides, homosteroids, norsteroids, and secosteroids.

Vitamin
A substance, usually an organic chemical complex, present in natural products or made synthetically, which is essential in the diet of man or other higher animals. Included here are vitamin precursors, provitamins, and vitamin supplements.

Classifications

Classification
A term or system of terms denoting an arrangement by class or category.

Conceptual Entity
A broad type for grouping abstract entities or concepts.

Functional Concept
A concept which is of interest because it pertains to the carrying out of a process or activity.

Group Attribute
A conceptual entity which refers to the frequency or distribution of certain characteristics or phenomena in certain groups.

Idea or Concept
An abstract concept, such as a social, religious or philosophical concept.

Intellectual Product
A conceptual entity resulting from human endeavor. Concepts assigned to this type generally refer to information created by humans for some purpose.
Language
The system of communication used by a particular nation or people.

Qualitative Concept
A concept which is an assessment of some quality, rather than a direct measurement.

Quantitative Concept
A concept which involves the dimensions, quantity or capacity of something using some unit of measure, or which involves the quantitative comparison of entities.

Regulation or Law
An intellectual product resulting from legislative or regulatory activity.

Spatial Concept
A location, region, or space, generally having definite boundaries.

Temporal Concept
A concept which pertains to time or duration.

Devices

Drug Delivery Device
A medical device that contains a clinical drug or drugs.

Medical Device
A manufactured object used primarily in the diagnosis, treatment, or prevention of physiologic or anatomic disorders.

Research Device
A manufactured object used primarily in carrying out scientific research or experimentation.

Disorders

Acquired Abnormality
An abnormal structure, or one that is abnormal in size or location, found in or deriving from a previously normal structure. Acquired abnormalities are distinguished from diseases even though they may result in pathological functioning (e.g., "hernias incarcerate").
Anatomical Abnormality
An abnormal structure, or one that is abnormal in size or location.

Cell or Molecular Dysfunction
A pathologic function inherent to cells, parts of cells, or molecules.

Congenital Abnormality
An abnormal structure, or one that is abnormal in size or location, present at birth or evolving over time as a result of a defect in embryogenesis.

Disease or Syndrome
A condition which alters or interferes with a normal process, state, or activity of an organism. It is usually characterized by the abnormal functioning of one or more of the host's systems, parts, or organs. Included here is a complex of symptoms descriptive of a disorder.

Experimental Model of Disease
A representation in a non-human organism of a human disease for the purpose of research into its mechanism or treatment.

Finding
That which is discovered by direct observation or measurement of an organism attribute or condition, including the clinical history of the patient. The history of the presence of a disease is a 'Finding' and is distinguished from the disease itself.

Injury or Poisoning
A traumatic wound, injury, or poisoning caused by an external agent or force.

Mental or Behavioral Dysfunction
A clinically significant dysfunction whose major manifestation is behavioral or psychological. These dysfunctions may have identified or presumed biological etiologies or manifestations.

Neoplastic Process
A new and abnormal growth of tissue in which the growth is uncontrolled and progressive. The growths may be malignant or benign.

Pathologic Function
A disordered process, activity, or state of the organism as a whole, of a body system or systems, or of multiple organs or tissues. Included here are normal responses to a negative stimulus as well as pathologic conditions or states that are less specific than a disease. Pathologic functions frequently have systemic effects.
Sign or Symptom
An observable manifestation of a disease or condition based on clinical judgment, or a manifestation of a disease or condition which is experienced by the patient and reported as a subjective observation.

Genes and Molecular Sequences

Amino Acid Sequence
The sequence of amino acids as arrayed in chains, sheets, etc., within the protein molecule. It is of fundamental importance in determining protein structure.

Carbohydrate Sequence
The sequence of carbohydrates within polysaccharides, glycoproteins, and glycolipids.

Gene or Genome
A specific sequence, or in the case of the genome the complete sequence, of nucleotides along a molecule of DNA or RNA (in the case of some viruses) which represent the functional units of heredity.

Molecular Sequence
A broad type for grouping the collected sequences of amino acids, carbohydrates, and nucleotide sequences. Descriptions of these sequences are generally reported in the published literature and/or are deposited in and maintained by databanks such as GenBank, European Molecular Biology Laboratory (EMBL), National Biomedical Research Foundation (NBRF), or other sequence repositories.

Nucleotide Sequence
The sequence of purines and pyrimidines in nucleic acids and polynucleotides. Included here are nucleotide-rich regions, conserved sequence, and DNA transforming region.

Geographic Areas

Geographic Area
A geographic location, generally having definite boundaries.

Living Beings

Age Group
An individual or individuals classified according to their age.
Alga
A chiefly aquatic plant that contains chlorophyll, but does not form embryos during development and lacks vascular tissue.

Amphibian
A cold-blooded, smooth-skinned vertebrate which characteristically hatches as an aquatic larva, breathing by gills. When mature, the amphibian breathes with lungs.

Animal
An organism with eukaryotic cells, and lacking stiff cell walls, plastids and photosynthetic pigments.

Archaeon
A member of one of the three domains of life, formerly called Archaebacteria under the taxon Bacteria, but now considered separate and distinct. Archaea are characterized by: 1) the presence of characteristic tRNAs and ribosomal RNAs; 2) the absence of peptidoglycan cell walls; 3) the presence of ether-linked lipids built from branched-chain subunits; and 4) their occurrence in unusual habitats. While archaea resemble bacteria in morphology and genomic organization, they resemble eukarya in their method of genomic replication.

Bacterium
A small, typically one-celled, prokaryotic micro-organism.

Bird
A vertebrate having a constant body temperature and characterized by the presence of feathers.

Family Group
An individual or individuals classified according to their family relationships or relative position in the family unit.

Fish
A cold-blooded aquatic vertebrate characterized by fins and breathing by gills. Included here are fishes having either a bony skeleton, such as a perch, or a cartilaginous skeleton, such as a shark, or those lacking a jaw, such as a lamprey or hagfish.

Fungus
A eukaryotic organism characterized by the absence of chlorophyll and the presence of a rigid cell wall. Included here are both slime molds and true fungi such as yeasts, molds, mildews, and mushrooms.

Group
A conceptual entity referring to the classification of individuals according to certain shared characteristics.
**Human**
Modern man, the only remaining species of the Homo genus.

**Invertebrate**
An animal which has no spinal column.

**Mammal**
A vertebrate having a constant body temperature and characterized by the presence of hair, mammary glands and sweat glands.

**Organism**
Generally, a living individual, including all plants and animals.

**Patient or Disabled Group**
An individual or individuals classified according to a disability, disease, condition or treatment.

**Plant**
An organism having cellulose cell walls, growing by synthesis of inorganic substances, generally distinguished by the presence of chlorophyll, and lacking the power of locomotion. Plant parts are included here as well.

**Population Group**
An individual or individuals classified according to their sex, racial origin, religion, common place of living, financial or social status, or some other cultural or behavioral attribute.

**Professional or Occupational Group**
An individual or individuals classified according to their vocation.

**Reptile**
A cold-blooded vertebrate having an external covering of scales or horny plates. Reptiles breathe by means of lungs and are generally egg-laying.

**Rickettsia or Chlamydia**
An organism intermediate in size and complexity between a virus and a bacterium, and which is parasitic within the cells of insects and ticks. Included here are all the chlamydias, also called "PLT" for psittacosis- lymphogranuloma venereum-trachoma.

**Vertebrate**
An animal which has a spinal column.
Virus
An organism consisting of a core of a single nucleic acid enclosed in a protective coat of protein. A virus may replicate only inside a host living cell. A virus exhibits some but not all of the usual characteristics of living things.

Objects

Entity
A broad type for grouping activities, processes and states.

Food
Any substance generally containing nutrients, such as carbohydrates, proteins, and fats, that can be ingested by a living organism and metabolized into energy and body tissue. Some foods are naturally occurring, others are either partially or entirely made by humans.

Manufactured Object
A physical object made by human beings.

Physical Object
An object perceptible to the sense of vision or touch.

Substance
A material with definite or fairly definite chemical composition.

Occupations

Biomedical Occupation or Discipline
A vocation, academic discipline, or field of study related to biomedicine.

Occupation or Discipline
A vocation, academic discipline, or field of study, or a subpart of an occupation or discipline.

Organizations
Health Care Related Organization
An established organization which carries out specific functions related to health care delivery or research in the life sciences.

Organization
The result of uniting for a common purpose or function. The continued existence of an organization is not dependent on any of its members, its location, or particular facility. Components or subparts of organizations are also included here. Although the names of organizations are sometimes used to refer to the buildings in which they reside, they are not inherently physical in nature.

Professional Society
An organization uniting those who have a common vocation or who are involved with a common field of study.

Self-help or Relief Organization
An organization whose purpose and function is to provide assistance to the needy or to offer support to those sharing similar problems.

Phenomena

Biologic Function
A state, activity or process of the body or one of its systems or parts.

Environmental Effect of Humans
A change in the natural environment that is a result of the activities of human beings.

Human-caused Phenomenon or Process
A phenomenon or process that is a result of the activities of human beings.

Laboratory or Test Result
The outcome of a specific test to measure an attribute or to determine the presence, absence, or degree of a condition.

Natural Phenomenon or Process
A phenomenon or process that occurs irrespective of the activities of human beings.

Phenomenon or Process
A process or state which occurs naturally or as a result of an activity.
Physiology

Cell Function
A physiologic function inherent to cells or cell components.

Clinical Attribute
An observable or measurable property or state of an organism of clinical interest.

Genetic Function
Functions of or related to the maintenance, translation or expression of the genetic material.

Mental Process
A physiologic function involving the mind or cognitive processing.

Molecular Function
A physiologic function occurring at the molecular level.

Organ or Tissue Function
A physiologic function of a particular organ, organ system, or tissue.

Organism Attribute
A property of the organism or its major parts.

Organism Function
A physiologic function of the organism as a whole, of multiple organ systems, or of multiple organs or tissues.

Physiologic Function
A normal process, activity, or state of the body.

Procedures

Diagnostic Procedure
A procedure, method, or technique used to determine the nature or identity of a disease or disorder. This excludes procedures which are primarily carried out on specimens in a laboratory.
Educational Activity
An activity related to the organization and provision of education.

Health Care Activity
An activity of or relating to the practice of medicine or involving the care of patients.

Laboratory Procedure
A procedure, method, or technique used to determine the composition, quantity, or concentration of a specimen, and which is carried out in a clinical laboratory. Included here are procedures which measure the times and rates of reactions.

Molecular Biology Research Technique
Any of the techniques used in the study of or the directed modification of the gene complement of a living organism.

Research Activity
An activity carried out as part of research or experimentation.

Therapeutic or Preventive Procedure
A procedure, method, or technique designed to prevent a disease or a disorder, or to improve physical function, or used in the process of treating a disease or injury.

Download the Predication Database

You can also download the predication database in the following way.

1) Use the ftp client and go to ftp://lhcftp.nlm.nih.gov
2) For user prompt, type "anonymous" and just type return for the password
3) Once you logon, go to the directory, anonymous/outgoing/cgsb
4) You will see a gzip file "semmedVER21_to01312012_02132012.sql.gz" and download it
5) Once you download it, you can unzip it and restore the database with the following command
   a. % Mysql --user=root --password semmedVER21 < semmedVER21_to01312012_02132012.sql
   b. FYI, you can install Windows version of MySQL as well as Linux or Solaris version

Here is also attached the description of the predication database. It has 56 million predications obtained from 20.7 million PubMed citations.

Description of the Predication Database

Documentation (Word)

Description of Semantic Medline Database
Date: 02/02/2012
Tables

Name: CONCEPT
Brief description: the table that has UMLS concepts
Column:
CONCEPT_ID: Auto generated primary key for each concept
CUI: CUI of the concept in UMLS
TYPE: If the concept comes from UMLS, it is “META”. If the concept comes from ENTREZ gene it is “ENTREZ”.
PREFERRED_NAME: The preferred name of the concept in UMLS.
GHR: The corresponding GHR name
OMIN: The corresponding OMIN reference

Name: CONCEPT_SEMTYPE
Brief description: The table that has semantic type for each concept in CONCEPT table
Column:
CONCEPT_SEMTYPE_ID: Auto generated primary key for each semtype of a concept
CONCEPT_ID: The foreign key to the CONCEPT
SEMTYPE: four character semantic type like “aapp” and “gngm”
NOVEL: Either “Y” (Yes) or “N” (No) representing the concept is novel or not. A novel concept is not too much broader and thus not too much close to root concept in UMLS.

Name: CONCEPT_TRANSLATION
Brief description: The table that has translated concepts in foreign languages for concepts in CONCEPT table
Column:
CONCEPT_TRANSLATION_ID: Auto generated primary key for each concept
CONCEPT_ID: The foreign key to the CONCEPT
LANGUAGE: It is three characters and represents foreign language. For example, “GER” means German.
TRANSLATION: Translated concept in the target language

Name: PREDICATION
Brief description: The table that keeps all the RELATION parts of the predications
Columns:
PREDICATION_ID: Auto generated primary key for each RELATION
PREDICATE: The string representation of each relation (for example PROCESS_OF)
TYPE: It can be ignored since we only use “semrep” in this application

Name: PREDICATION_ARGUMENT
Brief description: The table that keeps all the argument parts of the predications
Column:
PREDICATION_ARGUMENT_ID: Auto generated primary key for each argument of a predication
PREDICATION_ID: The foreign key of the relation in PREDICATION table that is used with this argument
CONCEPT_SEMTYPE_ID: The foreign key of semantic type of the concept semantic type. The semantic type information comes from each relation in the semrep output.
TYPE: a character representing the argument is used as the subject (“S”) or the object (“O”).
Name: SENTENCE

Brief description: The sentence information

Column:
- SENTENCE_ID: Auto generated primary key for each sentence
- PMID: The PMID of each sentence
- TYPE: This is either “ti” (representing title), or “ab” (abstract).
- NUMBER: This represents the location of the sentence either in title or paragraph.
- SENTENCE: The actual string of this sentence.

Name: SENTENCE_PREDICATION

Brief description: The mapping with sentence and derived predications

Column:
- SENTENCE_PREDICATION_ID: Auto generated primary key for each map of sentence and predication
- SENTENCE_ID: The foreign key to the sentence
- PREDICATION_ID: The foreign key to the PREDICATION of the derived predication
- PREDICATION_NUMBER: The number of the same predication that occurs in the same sentence. For instance, the same predication can occur more than once in the same sentence. If so, the PREDICATION_NUMBER of the first predication is 1 and the second is 2 and so on.
- Other columns: Auxiliary information that are not used in Semantic Medline

Name: PREDICATE_AGGREGATE

Brief information: The aggregate table that is obtained by joining the above tables and readily available to the SemMed Web application. SemMed looks at this table for retrieving predications, but also looks at the above tables to generate graphs in XML format to be visualized in Adobe Flash. The SQL script to get this table is:

```sql
insert ignore into PREDICATE_AGGREGATE (PID,SID,PNUMBER, PMID,predicate,s_cui,s_name,s_type,s_novel,o_cui,o_name,o_type,o_novel) select P.PREDICATION_ID, S.SENTENCE_ID, SP.PREDICATION_NUMBER, S.PMID, P.PREDICATE, group_concat(SC.CUI separator '|||'), group_concat(SC.PREFERRED_NAME separator '|||'),group_concat(SS.SEMTYPE separator '|||'),max(SS.NOVEL)="Y",group_concat(OC.CUI separator '|||'), group_concat(OC.PREFERRED_NAME separator '|||'),group_concat(OS.SEMTYPE separator '|||'),max(OS.NOVEL)="Y" from (select * from SENTENCE) S, SENTENCE_PREDICATION SP, PREDICATION P, CONCEPT SC, CONCEPT_SEMTYPE SS, PREDICATION_ARGUMENT SPA, CONCEPT OC, CONCEPT_SEMTYPE OS, PREDICATION_ARGUMENT OPA
WHERE S.SENTENCE_ID = SP.SENTENCE_ID and SP.PREDICATION_ID = P.PREDICATION_ID and SPA.PREDICATION_ID = P.PREDICATION_ID and SPA.TYPE="S" AND SS.CONCEPT_SEMTYPE_ID = SPA.CONCEPT_SEMTYPE_ID AND SC.CONCEPT_ID = SS.CONCEPT_ID AND OPA.PREDICATION_ID = P.PREDICATION_ID and OPA.TYPE="O" AND OS.CONCEPT_SEMTYPE_ID = OPA.CONCEPT_SEMTYPE_ID AND OC.CONCEPT_ID = OS.CONCEPT_ID group by SP.PREDICATION_ID, SP.SENTENCE_ID, SP.PREDICATION_NUMBER
```

Table Relation

CONCEPT to CONCEPT_SEMTYPE is 1-to-many relation through CONCEPT_ID sincere a concept can have more than 1 semantic type.
For example, a concept “beta beta 2-glycoprotein I” has two semantic type “aapp” and “bacs” and thus has two entries in CONCEPT_SEMTYPE as shown below.
PREDICATION to PREDICATION_ARGUMENT is 1-to-many since a predication has subject and object. Look at the following example. It shows that the same predication ID appears twice in PREDICATION_ARGUMENT table.

<table>
<thead>
<tr>
<th>PREDICATION_ARGUMENT_ID</th>
<th>PREDICATION_ID</th>
<th>CONCEPT_SEMTYPE_ID</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>15450</td>
<td>771545</td>
<td>121618</td>
<td>O</td>
</tr>
<tr>
<td>15450</td>
<td>671545</td>
<td>497358</td>
<td>S</td>
</tr>
</tbody>
</table>

SENTENCE to SENTENCE_PREDICATION is 1-to-many since more than a predication appears in a sentence. Look at the following example. It shows that the same SENTENCE ID appears MANY TIMES in SENTENCE_PREDICATION table.

<table>
<thead>
<tr>
<th>SENTENCE_PREDICATION_ID</th>
<th>SENTENCE_ID</th>
<th>PREDICATION_ID</th>
<th>PREDICATION_NUMBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>92736</td>
<td>185267</td>
<td>71621</td>
<td>1</td>
</tr>
<tr>
<td>92737</td>
<td>185267</td>
<td>71622</td>
<td>1</td>
</tr>
<tr>
<td>92738</td>
<td>185267</td>
<td>71623</td>
<td>1</td>
</tr>
<tr>
<td>92739</td>
<td>185267</td>
<td>71623</td>
<td>2</td>
</tr>
<tr>
<td>92740</td>
<td>185267</td>
<td>71624</td>
<td>1</td>
</tr>
<tr>
<td>92741</td>
<td>185267</td>
<td>71624</td>
<td>2</td>
</tr>
</tbody>
</table>

**Predication Databases**


Index of /outgoing/cgsb/

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline">Link to the index</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Updated: Wed, 23 Sep 2015 04:23:19 GMT

Powered by Qmindtouch™
My SQL


Source: https://www.mysql.com/

Semantic MEDLINE: Semantic Predications for Biomedical Research
Introduction

- Augment document retrieval systems
- Manipulate information - Not just documents
- Bridge the gap between
  - Language (text)
  - Meaning
- Support emerging applications
  - Text mining (task-driven extraction of facts)
  - Question answering
  - Literature-based discovery
Semantic MEDLINE

SemRep: Extract all predications

Aromatase inhibitors, Malignant neoplasm of breast, Tamoxifen, Malignant neoplasm of breast, CDKN1A gene, Malignant neoplasm of breast, BARD1 gene, Tamoxifen, Patients, Malignant neoplasm of breast, Individual
Conduct research

Semantic Processing for Managing the Biomedical Research Literature

Acknowledgments

- Marcelo Fiszman, M.D., Ph.D.
- Halil Kilicoglu, Ph.D.
- Graciela Roseblat, Ph.D.
- Dongwook Shin, Ph.D.
- Christopher M. Miller, M.D.
- Michael J. Cairelli, D.O.
- Guocal Chen, Ph.D.
- T. Elizabeth Workman, Ph.D.
- Susan Roy, M.S., M.L.S.
- J. Caleb Goodwin
- Bartłomiej Wilkowski, Ph.D.
Semantic Processing for Managing the Biomedical Research Literature

Access to online text

- Document retrieval systems
  - Google
  - PubMed for biomedical information
- Technology: Manipulate text strings
  - Frequency of occurrence
  - Distribution patterns
Access to online text

- Document retrieval systems
  - Google
  - PubMed for biomedical information
- Technology: Manipulate text strings
  - Frequency of occurrence
  - Distribution patterns
  - No access to meaning

Emerging applications

- Text mining
  - Task-driven extraction of facts
  - Observe trends
- Connect text and structured data
- Automatic summarization
- Question answering
- Literature-based discovery
  - Research assistance
Emerging applications

- Text mining
  - Task-driven extraction of facts
  - Observe trends
- Connect text and structured data
- Automatic summarization
- Question answering
- Literature-based discovery
  - Research assistance
- Require more effective language processing

Automatic semantic interpretation

- Augment document retrieval systems
- Manipulate information
  - Not just documents
- Bridge the gap between
  - Language (text)
  - Meaning
- Summarize and visualize information

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by MindTouch
Automatic semantic interpretation

- Augment document retrieval systems
- Manipulate information
  - Not just documents
- Bridge the gap between
  - Language (text)
  - Meaning
- Summarize and visualize information
- **Semantic MEDLINE**

Semantic MEDLINE

- PubMed
- MEDLINE citations
- SemRep
- Semantic predications
- Automatic summarization
- Graphical summary
- Enhanced access to biomedical literature
Semantic MEDLINE

SemRep: Extract predication

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
SemRep: Extract predication 2

... Exemestane after non-steroidal aromatase inhibitor for post-menopausal women with advanced breast cancer

Aromatase Inhibitor \texttt{TREATS} Malignant neoplasm of breast

Comment: Provides a conceptual overview of the contents of the citations returned

SemRep: Extract all predications

Abstraction summarization

Comment: Provides a conceptual overview of the contents of the citations returned
**Abstraction summarization**

- Specify a topic
- Retain predications on the topic using a schema
- Eliminate uninformative predications
- Retain most frequent predications

---

**Summarized predications**

![Diagram of summarized predications]

- Aromatase Inhibitors TREATS Breast carcinoma
- Tamoxifen TREATS Breast carcinoma
- CDKN1A gene ASSOCIATED_WITH Breast carcinoma
- CDKN1A gene STIMULATES BARD1 gene
- Tamoxifen TREATS Patients
- Malignant neoplasm of breast TREATS Individual
Conduct research

An alpha-fetoprotein-derived peptide reduces the uterine hyperplasia and increases the antihormone effect of tamoxifen.

Andersen TT, Georgekutty J, Defreeze LA, Amaratunga G, Naandiran A, Lemoniski L, Jacobsson H, Bennett IA. Center for Cardiovascular Sciences, Albany Medical College, Albany, NY 12208, USA. anderss@maax.amc.edu

Tamoxifen (Tam) is effective for the treatment and prevention of breast cancer. However, it has toxic side effects and has limited duration utility because, over time, human tumors become resistant to Tam. Recently, a new recombinant peptide, alpha-fetoprotein-derived peptide (AFH1), has been proposed for the treatment and prevention of breast cancer. The purpose of this paper is to determine whether combining AFH1 with Tam would increase efficacy and reduce toxicity in experimental models of breast cancer. Low doses of AFH1 and Tam were more effective in combination than either agent alone against breast cancer growth in cell culture, in tumour xenografts in mature mice, and in carcinogen-exposed rats. Alpha-Fetoprotein-derived peptide interfered with Tam-induced uterine hyperplasia in immature mice, and showed no

NLP technology: SemRep

Comment: Both language and human thought are large, for feasibility we need to scale down the complexity of the process of semantic interpretation.
NLP technology: SemRep

- Abstract away from full interpretation
  - Use underspecified syntactic structure
  - Limit to core semantic predicates
  - Focus interpretation by domain
- Guided by linguistic principles
  - Generalizations about the structure of English
- Accommodate incremental development
- Exploit structured domain knowledge
  - Unified Medical Language System (modified)

Components of the UMLS

- SPECIALIST Lexicon
  - More than 432,822 entries (general and medical)
  - Syntactic information (e.g., plurals and verbal inflection)
- Metathesaurus
  - 2,181,676 concepts containing variant terms with synonymous meaning
  - Concepts are assigned one or more semantic types
- Semantic Network
  - 135 semantic types (or categories)
  - 54 relationships
Metathesaurus concept

- Concept name
  - Arthroplasty
- Synonyms
  - Reconstruction of joint
  - Repair of joint ...
- Semantic type
  - Therapeutic or Preventive Procedure

Semantic Network 1
Semantic Network Relationships

Pharmacologic Substance TREATS Disease or Syndrome
Therapeutic or Preventive Procedure USES Medical Device
Body Location or Region LOCATION_OF Biologic Function
Disease or Syndrome OCCURS_IN Population Group
Disease or Syndrome PROCESS_OF Organism

Semantic phenomena 1

- Concentrate on propositions

The results provide strong evidence that **imipramine treats panic disorder**
Semantic phenomena

- Concentrate on propositions
- Do not address
  - Attitudes

The results provide strong evidence that imipramine treats panic disorder

Related research in biomedicine

- MedLEE, GENIES
  - Semantic grammar
- AQUA
  - Definite clause grammar
- MPLUS
  - Chart parser
- MEDSYNDIKATE
  - Dependency grammar

Updated: Wed, 23 Sep 2015 04:23:19 GMT
 Powered by mindtouch
Processing overview

- Syntactic analysis
  - Lexical look-up
  - Tagging
  - Underspecified parser (chunker)
- Concept recognition
  - MetaMap
  - Special processing for genes and proteins
- Construct predication
  - Indicator rules map to Semantic Network
  - Syntactic constraints
  - Semantic constraints

Input text

aggressive combination chemotherapy in the management of hypercaltorrine renal failure
Lexical look-up and tagger

Parser
SemRep 3

SemRep 4

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
SemRep predication database

- MEDLINE (National Library of Medicine)
  - Bibliographic database of the biomedical research literature
  - More than 21 million citations (1940s to present)
- SemRep processing
  - All of MEDLINE
  - 56 million predications (26 predicate types)
- Made available to the research community
  - SQL database
  - RDF triples (Olivier Bodenreider)
Several evaluations

- Focused on biomedical subdomains, e.g.
  - Clinical treatment
  - Genetic etiology of disease
  - Pharmacogenomics
- Focused on structure, e.g.
  - Hypenymic propositions
  - Comparatives
  - Nominalizations
- Precision is around 75% (lower for molecular biology)
- Recall is around 60%

Expand domain coverage

- Originally for clinical medicine and basic biomedical research
- Extended to influenza epidemic preparedness
- Currently working on
  - Health promotion (Melissa Resnick)
  - Climate and health (Sally Howe, Benjamin Brookstone)
  - Biomedical information management
  - Expanding coverage of basic biology research
- General methodology for exploiting existing ontologies (Graciela Roseblat)
A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men

Source: http://www.biograph.be/concept/graph/C0020619/C0039601

Abstract

**STUDY OBJECTIVES:**

Sleep quality commonly diminishes with age, and, further, aging men often exhibit a wider range of sleep pathologies than women. We used a freely available, web-based discovery technique (Semantic MEDLINE) supported by semantic relationships to automatically extract information from MEDLINE titles and abstracts.

**DESIGN:**

We assumed that testosterone is associated with sleep (the A-C relationship in the paradigm) and looked for a mechanism to explain this association (B explanatory link) as a potential or partial mechanism underpinning the etiology of eroded sleep quality in aging men.

**MEASUREMENTS AND RESULTS:**

Review of full-text papers in critical nodes discovered in this manner resulted in the proposal that testosterone enhances sleep by inhibiting cortisol. Using this discovery method, we posit, and could confirm as a novel hypothesis, cortisol as part of a mechanistic link elucidating the observed correlation between decreased testosterone in aging men and diminished sleep quality.
CONCLUSIONS:

This approach is publically available and useful not only in this manner but also to generate from the literature alternative explanatory models for observed experimental results.

PMID:

22294819

[PubMed - in process]

PMCID:

PMC3250368

[Available on 2012/8/1]

Semantic Medline: Multi-Document Summarization and Visualization

Source: http://www.nlm.nih.gov/pubs/techbull...t/semantic.ppt

Title Page
Information management application

- Manipulate information in addition to documents
- Manage the results of PubMed searches
  - Help users decide what to read
- Connect knowledge from various sources
- Integrate application interfaces

Managing retrieval results

- Information retrieval
  - 500 citations

Semantic Medline

Network of relations

Updated: Wed, 23 Sep 2015 04:23:19 GMT
Managing retrieval results 2

Seamless integration of technologies

- Information retrieval
  - PubMed - MEDLINE
- Natural language processing: SemRep
  - Represent content of text with semantic predicates
- Abstraction summarization
  - Informative: Overview of most salient information
- Visualization
  - Indicative: Links to source text and additional information
Semantic Medline Overview

Document selection

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
Semantic interpretation

... aromatase inhibitor provides mortality benefit in early breast carcinoma ...

Aromatase Inhibitors \(\text{treats}\) Breast Carcinoma

... determined the spectrum and frequency of ATM missense variants in 443 breast cancer patients ...

ATM gene \(\text{associated with}\) Breast Carcinoma

Semantic predications

Tamoxifen \(\text{treats}\) Breast Carcinoma

Aromatase Inhibitors \(\text{treats}\) Breast Carcinoma

ATM gene \(\text{associated with}\) Breast Carcinoma

Tamoxifen \(\text{treats}\) Patients

Breast Carcinoma \(\text{process of}\) Individual
Summarization

Abstraction summarization

- Specify a topic
- Retain predications on the topic
- Eliminate uninformative predications
- Retain most frequent predications
Salient semantic predications

Visualization
Informative graph

Semantic Medline Live
Summarization options

Summarization results
More detail

Substance interactions
Semantic MEDLINE: An advanced information management application for biomedicine

Source: http://iospress.metapress.com/content/u2j7035q26h5i421/fulltext.html

Information Services & Use 31 (2011) 15–21
DOI 10.3233/ISU-2011-0627
IOS Press

Semantic MEDLINE: An advanced information management application for biomedicine
Thomas C. Rindflesch*, Halil Kilicoglu, Marcelo Fiszman, Graciela Rosemblat and Dongwook Shin
Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health,
Bethesda, MD, USA

* Corresponding author: Thomas C. Rindflesch, PhD, Lister Hill National Center for Biomedical Communications,
National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894, USA. Tel.: +1 301 435 3191; Fax: +1 301 496 0673; E-mail: tcr@nlm.nih.gov.

Abstract
To support more effective biomedical information management, Semantic MEDLINE integrates document retrieval, advanced natural language processing, automatic summarization and visualization into a single Web portal. The application is intended to help manage the results of PubMed searches by condensing core semantic content in the citations retrieved. Output is presented as a connected graph of semantic relations, with links to the original MEDLINE citations. The ability to connect salient information across documents helps users keep up with the research literature and discover connections which might otherwise go unnoticed. Semantic MEDLINE can make an impact on biomedicine by supporting scientific discovery and the timely translation of insights from basic research into advances in clinical practice and patient care. Semantic MEDLINE is illustrated here with recent research on the clock genes.
1 Introduction

Access to online text is provided by document retrieval systems such as Google and PubMed (for biomedical information). The underlying technology of these systems typically manipulates text strings, perhaps augmented by frequency of occurrence and distribution patterns, and has remained largely unchanged since the 1980s [21]. Such systems have no access to the meaning of the text being processed. In biomedical information management, for example, more effective language processing is needed to support emerging applications, such as text mining aimed at task-driven extraction of facts to observe trends [2], those that connect text and structured data [8], question answering systems [4] and literature based discovery [24], which provides assistance to scientific research. Automatic semantic interpretation (e.g., [7]) is intended to augment document retrieval systems by manipulating information, not just documents, and thereby bridge the gap between text and meaning. In the biomedical domain the Semantic MEDLINE application [14] exploits this technology to provide enhanced access to the research literature. The application calls on PubMed to return MEDLINE citations in response to a user’s query. Semantic processing extracts relationships representing the meaning of retrieved citations. After automatic summarization to focus on salient information, results are presented to the user as a connected graph of relationships for further exploration (Figure 1).
2 Semantic interpretation of biomedical text: SemRep

Semantic MEDLINE calls on the SemRep natural language processing system [18,19] to extract semantic relationships. SemRep inspects each sentence in input text and identifies some of the relationships representing the meaning of the sentence. Semantic relationships are represented as predications, a formal representation having a predicate and arguments. For example, in Figure 2, the predication “Genes AFFECTS Circadian Rhythms” was extracted from the sentence “Clock genes are the genes that control circadian rhythms in physiology and behavior”. SemRep semantic predications provide a normalized representation of (part of) the meaning of a sentence and can be further manipulated by computational means.

SemRep depends on domain knowledge in the Unified Medical Language system (UMLS) developed by the US National Library of Medicine [12]. The UMLS consists of three knowledge sources: the SPECIALIST Lexicon, Metathesaurus and Semantic Network [6]. SemRep exploits the latter two for domain knowledge. The Metathesaurus contains a very large number of concepts compiled from an array of knowledge sources in biomedicine, focused on the clinical domain. A Metathesaurus concept consists of a collection of synonyms drawn from constituent vocabularies and terminologies. One of these is selected as the concept name. Concepts cover areas such as diseases, physiologic processes, drugs, anatomical concepts, organisms and population groups and concept classes that represent the same areas are grouped into semantic types, such as “Disease or Syndrome”, “Pharmacologic Substance” and “Population Group”.

http://semanticommunity.info/A_NITRD_Dashboard/Semantic_Medline
Updated: Wed, 23 Sep 2015 04:23:19 GMT
Powered by mindtouch
A SemRep predication has UMLS Metathesaurus concepts as arguments and a UMLS Semantic Network relation as predicate. In the example given earlier, both arguments (“Genes” and “Circadian Rhythms”) are Metathesaurus concepts, and the predicate (“AFFECTS”) is a relation in the Semantic Network that binds the two arguments in this instance. During processing, the underlined text is mapped to the arguments and the italicized verb is mapped to the predicate. The MetaMap program [3], is used to map text to Metathesaurus concepts. Each concept has a semantic type associated with it, as shown in the following example:

Concept name: Genes; Semantic type: Gene or Genome.
Concept name: Circadian Rhythms; Semantic type: Organism Function.

**Figure 2** Example predication extracted from text by SemRep

Top

**Figure 3** SemRep predications extracted from text on breast cancer. Predications outside the box are eliminated by automatic summarization

Bottom

The UMLS Semantic Network stipulates relationships between concepts, stated in terms of the semantic types assigned to Metathesaurus concepts. These relationships form a pattern for the semantic predications that SemRep is allowed to identify and extract from biomedical text. That is, the semantic types of the Metathesaurus concepts serving as
argument of SemRep predications must match those for the specific relationships in the Semantic Network, such as one of the following:

- Gene or Genome **AFFECTS** Organism Function.
- Therapeutic or Preventive Procedure **USES** Medical Device.
- Pharmacologic Substance **TREATS** Disease or Syndrome.
- Body Location or Region **LOCATION_OF** Biologic Function.
- Disease or Syndrome **OCCURS_IN** Population Group.
- Disease or Syndrome **PROCESS_OF** Organism.

SemRep extracts an array of predicate types in the biomedical domain. For example, the predications in Figure 3 were extracted from text on the topic of breast cancer, and provide information on various aspects of this disease. The predications inside the box above are more informative than those below. Automatic summarization \[9\] is used to eliminate the less informative predications. A major aspect of this processing is to use the UMLS to identify concepts that are too general to be informative, such as “Patient” and “Individual”. Predications having such concepts as arguments are eliminated.

Regarding domain coverage, SemRep was initially developed for clinical medicine (the major focus of the UMLS), and has been extended to genetic etiology and substance interactions \[20\], pharmacogenomics and molecular biology \[1\], as well as influenza epidemic preparedness \[13\]. Research currently addresses extending the program to documents on public health and climate and health. Consideration is being given to extending coverage beyond biomedicine, to biomedical informatics, which includes practical application in the computer science domain.

SemRep has been run on MEDLINE, a bibliographic database of the biomedical research literature compiled and maintained by the US National Library of Medicine, containing some 19 million citations (1940s to present). 7.6 million citations (dated 01/01/99 through 02/28/11) have been processed and more than 26 million semantic predications extracted. These are stored in an SQL database and RDF triple store, and made available to the research community. SemRep semantic predications support the Semantic MEDLINE application.

### 3. Semantic MEDLINE application

We illustrate the use of Semantic MEDLINE by searching on the clock genes, discovered in 1971 in fruit flies \[23\] and subsequently found in humans (and all organisms). Originally these genes were thought to control sleep-related circadian rhythms only \[16\]. However, recent research provides insight into their physiological consequences underpinning the etiology of a range of common disorders. In creating and exploiting a graphical summary in Semantic MEDLINE, the first step is to issue a PubMed query, in this case: "clock gene". We limit results to the 1000 most recent citations, and apply SemRep, which produces 404 semantic predications. These are summarized and displayed in a graph of nodes and arcs representing arguments and predicates (Figure 4). The nodes are color coded for UMLS semantic type. For example in Figure 4, diseases, such as "Metabolic Diseases" (about 10 o’clock in the central portion of the graph) are pink; physiologic functions (e.g., “Homeostasis” at 11 o’clock) are gray; and genes (“AANAT” at 1 o’clock) are mauve. Arcs representing predicates are also color coded. (Predications are read in the direction...
of the arrow.) For example cyan is used for the relation predisposes: “CLOCK PREDISPOSES Metabolic Diseases” (11 o’clock); magenta for INTERACTS_WITH: “CLOCK INTERACTS_WITH Phosphotransferases” (11 o’clock); and green for “AFFECTS “: “CLOCK AFFECTS Homeostasis” (12 o’clock).

The graph summarizes and organizes the content of the documents processed and can be exploited in two ways: by taking advantage of both its informative and indicative aspects. In summarization theory, “informative” refers to aspects of the summary itself, while “indicative” refers to the way in which additional information is provided.

Inspection of the nodes in Figure 4 reveals concepts central to research on the clock genes. For example, the major genes (CRY1, CRY2, PER1, PER2, DEC1, ARNTL, ARNTL2 and CLOCK gene) appear in the graph along with several disorders affected by them (tumor growth, Mood Disorders, Depressive symptoms, Metabolic syndrome, Winter Depression). Underlying physiologic functions (Physiological aspects, Circadian Rhythms, Growth) are seen as well. The relationships in which these concepts participate provide more specific information. For example, it is seen that CLOCK gene is ASSOCIATED_WITH (brown) Depressive symptoms (lower right section of the figure) and DISRUPTS (yellow) tumor growth (right-hand section of the figure).

**Figure 4 Semantic MEDLINE graph**

Showing SemRep predications extracted from MEDLINE citations retrieved with the PubMed query “clock gene”. (Colors are visible in the online version of the article; [http://dx.doi.org/10.3233/ISU-2011-0627](http://dx.doi.org/10.3233/ISU-2011-0627)) Note: In this version - [http://iospress.metapress.com/content/u2j7035q26h5l421/fulltext.html](http://iospress.metapress.com/content/u2j7035q26h5l421/fulltext.html)
More information is available by clicking on the arcs, which are linked to the MEDLINE citations from which the relationships were extracted. This facility allows the user to investigate relevant phenomena in more depth. For example, from selected relationships, three major aspects of the clock genes (as reflected in current research) can be determined:

1. The clock genes affect core aspects of physiology, especially metabolism.
2. The clock genes can be influenced by various factors, including body substances (e.g., hormones, cytokines), food and cognitive state.
3. Disruption of the clock genes can contribute to the development of metabolic disease.

The relevant relationships are labeled in Figure 4. Text in each of the abstracts from which the relationship was extracted illustrates the corresponding points noted.

• **CLOCK AFFECTS homeostasis** (labeled “1” in Figure 4) “The circadian clock controls energy homeostasis by regulating circadian expression of proteins involved in metabolism” [5].

• **CLOCK AFFECTS Body tissue** (labeled “2”) “This review will focus on the interconnection between the circadian clock and metabolism, with implications for obesity and how the circadian clock is influenced by hormones, nutrients, and timed meals” [10].

• **CLOCK PREDISPOSES Metabolic Diseases** (labeled “3”) “Emerging evidence suggests that circadian clock function is closely linked to metabolic homeostasis and that rhythm disruption can contribute to the development of metabolic disease” [15].
The Semantic MEDLINE graph supports the kind of patterns that underpin literature-based discovery [22]. One such pattern characterizes closed discovery, in which a relationship A → C can be explicated by finding two relationships, A → B and B → C, thereby providing B as a mechanistic explanation. As an example, we consider cancer and obesity. There has been a longstanding and pervasive research interest in the association between these two phenomena. The PubMed query “obesity AND (cancer OR neoplasm)” returns 11,091 citations. The earliest is from 1947 [17]. However, until recently, potential mechanisms involved have not been well understood. The clock genes constitute one such mechanism, which is partly explicated in Figure 4. We first look at the citation [11] from which the predication “CLOCK gene AFFECTS Metabolic syndrome” (labeled “4” in Figure 4) was extracted. The thrust of the research reported is that “. . . obesity induced by high-fat diet alters the circadian-clock system. . .” and further that “expressions of circadian-clock genes and circadian clock-controlled genes, including Per1–3. . . were altered in the livers and/or kidneys”.

With that in mind, we next inspect the predication “CLOCK gene DISRUPTS tumor growth” (labeled “5”) and the citation [25] from which it was generated. The citation contains two statements that implicate Per1 and Per2 in cancer: “Per2 is a core clock gene, the product of which suppresses cancer cell proliferation and tumor growth in vivo and in vitro”. And “Down-regulation of the expression of tumor Per1 increases cancer cell growth in vitro and tumor growth in vivo. . .”. Taken together, these two citations provide the clock genes as a mechanistic link (B) for the observed relationship between cancer and obesity (A → C). Although this is not an actual discovery, it illustrates the kind of methodology Semantic MEDLINE supports as a resource for biomedical researchers.

4. Conclusion

Emerging applications in biomedical information management require more expressive text analysis than that provided by available document retrieval systems. Semantic MEDLINE provides enhanced access to the biomedical research literature by combining PubMed document retrieval, semantic relationships, and automatic summarization. Results are presented to the user in a graph representing an overview of content with links to source documents. The core technology for this application is the SemRep natural language processing system, which extracts semantic information from biomedical text, supported by domain knowledge in the Unified Medical Language system. Semantic MEDLINE usage is illustrated with recent research literature on the clock genes, showing the value of this system for keeping abreast of research trends and suggesting the possibility of exploiting it for scientific discovery.

Acknowledgement

This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.
References


[21]  
D. Satterthwaite, Emerging technologies to speed information access, Information Services & Use 30 (2010), 99–105.

[22]  
D.R. Swanson, Fish oil, Raynaud’s syndrome, and undiscovered public knowledge, Perspectives in Biology and Medicine 30 (1986), 7–18.

[23]  

[24]  
M. Weeber, J.A. Kors and B. Mons, Online tools to support literature-based discovery in the life sciences, Briefings in Bioinformatics 6 (2005), 277–286.

[25]  