



Semantic MEDLINE: Knowledge Discovery and Hypothesis Generation from Biomedical Research Literature

Halil Kılıçoğlu

National Library of Medicine
National Institutes of Health

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.

Data to Knowledge in Biomedicine

- Biomedical “big data”
 - Clinical measurements, environmental data, phenotypic data, genomic data
 - Research articles, clinical notes
- Considerable effort on data infrastructure
 - Structured, curated information resources
 - Mechanisms and structures for data sharing, integration
- Less attention on how to use these resources
 - For knowledge discovery and hypothesis generation

Biomedical Research Literature

- Information overload
 - More than 26 million abstracts in PubMed
- Need for natural language processing (NLP)
 - Transform text into computable semantic representations
 - Allow aggregation at large scale
- Effective, automatic access to information
 - Biological database curation
 - Systematic review support

Overview

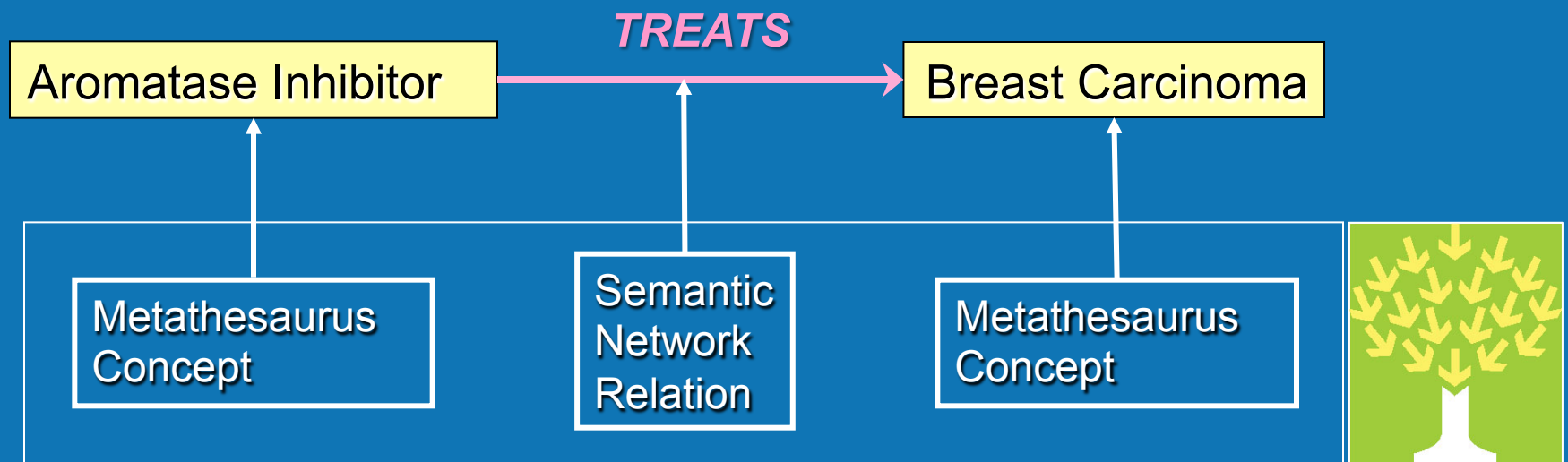
- SemRep
 - Relationship extraction from biomedical research articles
- Semantic MEDLINE
 - Advanced knowledge management, literature-based discovery, hypothesis generation

SemRep

- Extracts semantic predications
 - Subject-predicate-object triples
- Linguistically-oriented
- Relies on domain knowledge
 - Unified Medical Language System (UMLS)
 - Metathesaurus
 - Semantic Network
 - SPECIALIST Lexicon

Semantic Predications

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



Unified Medical Language System

SemRep Predicates

- **Clinical:** ADMINISTERED_TO, COMPLICATES, DIAGNOSES, PREDISPOSES, PREVENTS, OCCURS_IN, MANIFESTATION_OF, PROCESS_OF, TREATS
- **Molecular biology:** ASSOCIATED_WITH, AUGMENTS, CONVERTS_TO, DISRUPTS, INHIBITS, INTERACTS_WITH, STIMULATES
- **General:** AFFECTS, CAUSES, COEXISTS_WITH, ISA, LOCATION_OF, METHOD_OF, PART_OF, PRECEDES, PRODUCES, USES

SemRep Processing

- Syntactic analysis
 - Lexical look-up (SPECIALIST Lexicon)
 - Tagging
 - Underspecified parser (chunker)
- Concept recognition
 - MetaMap (map text to Metathesaurus)
 - Special processing for genes and proteins
- Predication construction
 - Indicator rules (map text to Semantic Network)
 - Syntactic and semantic constraints

SemRep Evaluation

- Focused on biomedical subdomains
 - Clinical treatment, genetic etiology of disease, pharmacogenomics
- Focused on linguistic structure
 - Hypernymic relations, comparatives, nominalizations
- Overall
 - Precision is around 75% (lower for molecular biology)
 - Recall is around 60%

SemMedDB

- Semantic predications from titles and abstracts of all PubMed articles
 - ~26M articles, ~91M predications (June 30, 2017)
- Updated biannually
- Made available to the research community
 - MySQL database
 - <http://skr3.nlm.nih.gov/SemMedDB/>

Semantic MEDLINE

- Helps navigate through the research literature
 - Combines document retrieval, semantic predications, summarization and network visualization
 - Make connections which might go unnoticed
- Literature-based discovery, hypothesis generation
 - Swanson's A-B-C discovery model
 - Discovery browsing
 - A-B-C-D-....

Semantic MEDLINE Overview

PubMed

MEDLINE abstracts

SemMedDB

Semantic predications

Automatic summarization

Graphical summary

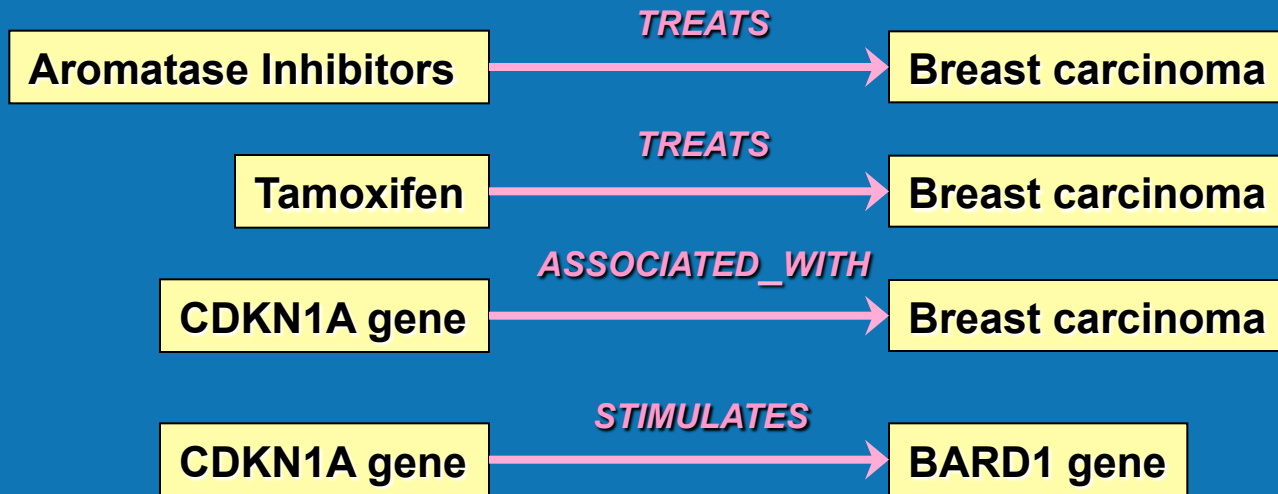
Biomedical information management

Summarization



- Specify a topic and summary view
- Retain predications on the topic using a schema (Relevance, Connectivity)
- Eliminate uninformative predications (Novelty)
- Retain most frequent predications (Saliency)

Summarized Predications



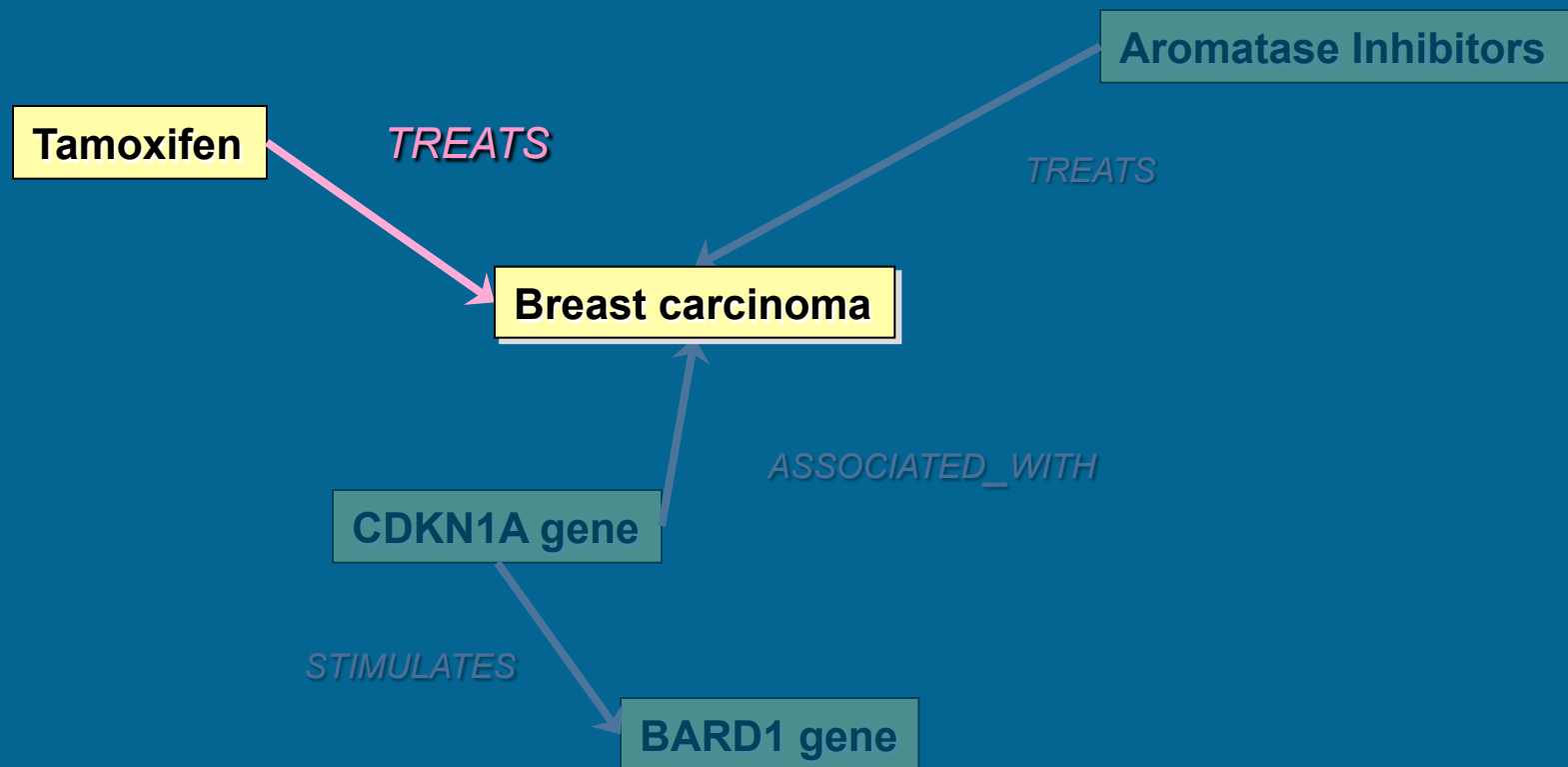
...

...

...



Semantic MEDLINE: Visualization

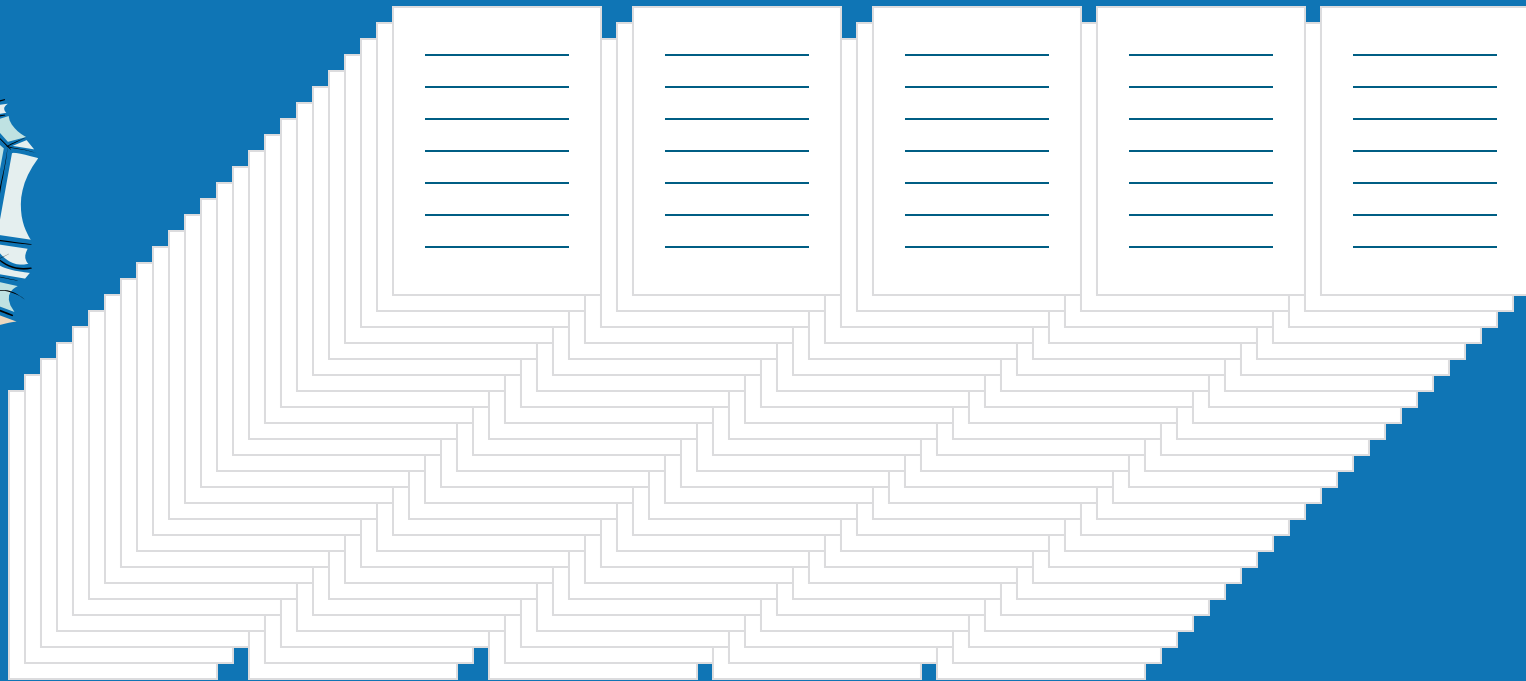


Semantic MEDLINE: Link to Text

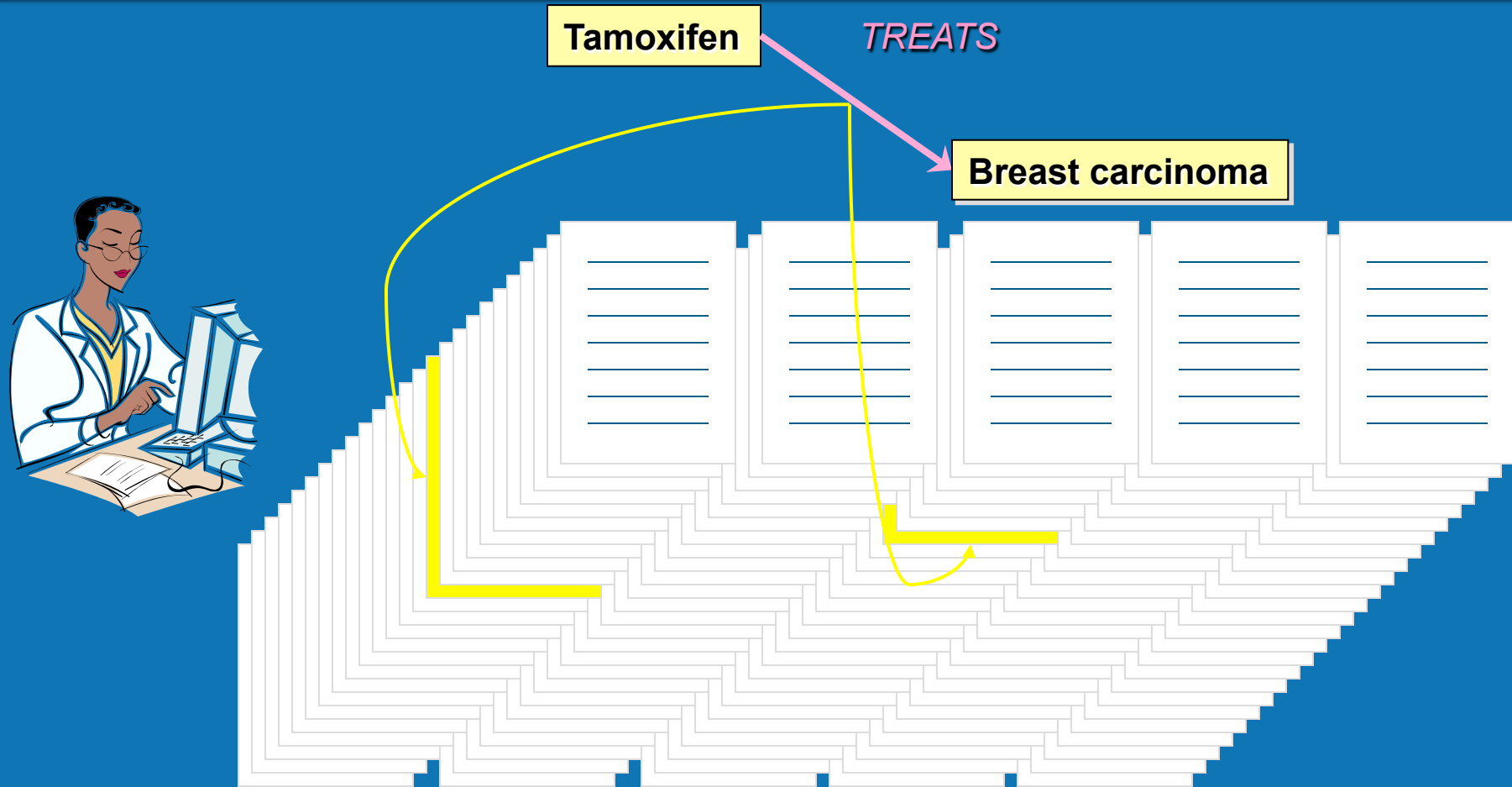
Tamoxifen

TREATS

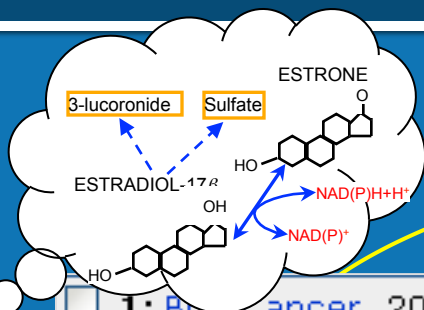
Breast carcinoma



Semantic MEDLINE: Link to Text



Semantic MEDLINE: Research



Tamoxifen

TREATS

Breast carcinoma

1: Breast Cancer, 2007 Aug 6;97(3):327-33. Epub 2007 Jul 17.

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

Andersen TT, Georgekutty J, Defreest LA, Amaratunga G, Narendran A, Lemanski N, Jacobson HI, Bennett JA.

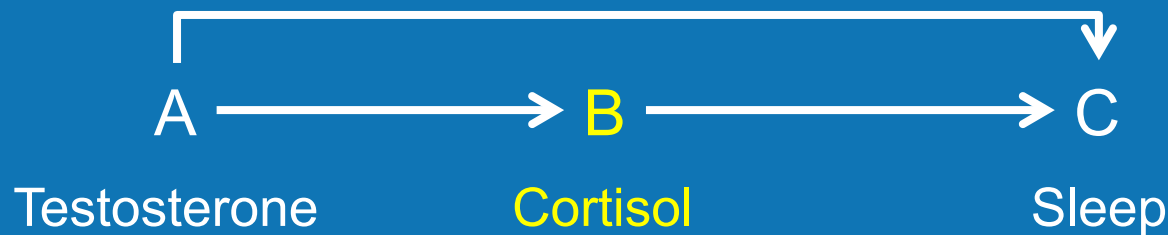
Center for Cardiovascular Sciences, Albany Medical College, Albany, NY 12208, USA.
anderst@mail.amc.edu

Tamoxifen (Tam) is effective for the treatment and prevention of breast cancer.

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

Semantic MEDLINE Applications

- Hypothesis generation
 - Closed discovery model
 - “Cortisol as part of a mechanistic link between decreased testosterone in aging men and diminished sleep quality.”



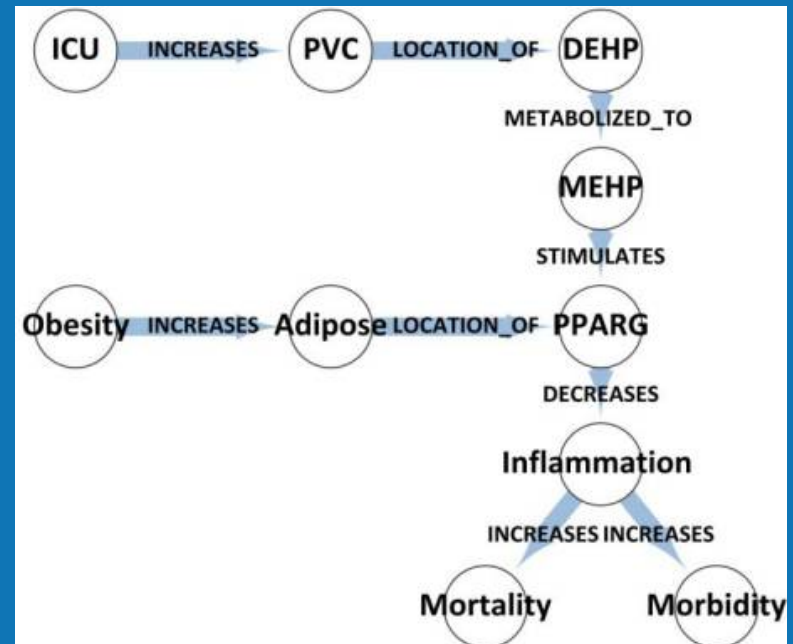
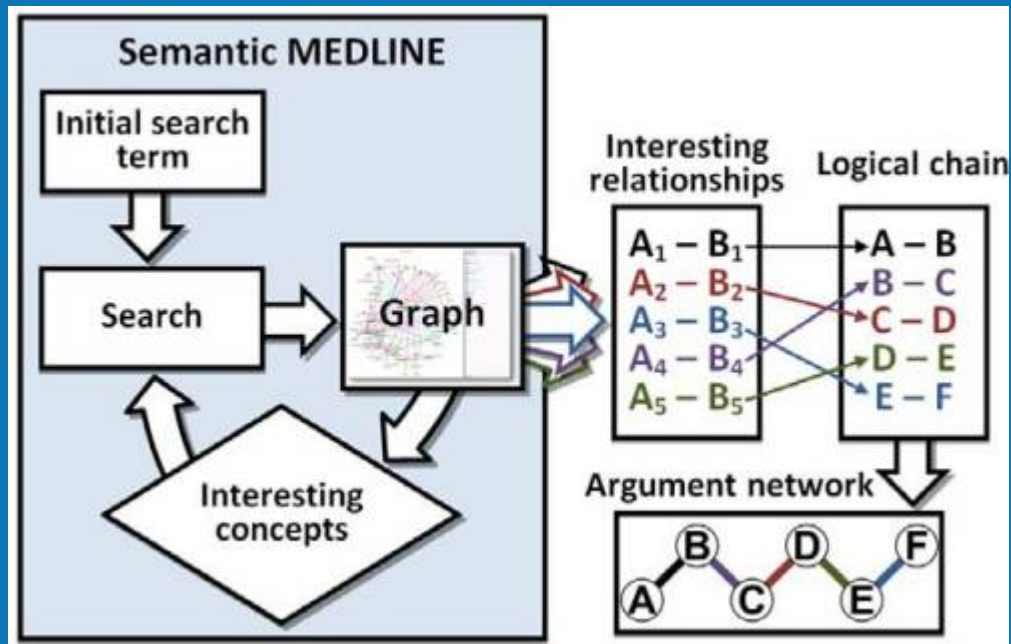
Semantic MEDLINE Applications

➤ Discovery browsing

- Increased obesity predicts decreased mortality and morbidity at ICU (“Obesity paradox”)
 - “PPAR- γ is greatly expressed in fat tissue and its activation is anti-inflammatory”
 - “DEHP activates PPAR- γ ”
 - “DEHP is commonly used in PVC and is leached from lines and bags at therapeutic doses in standard interventions in ICU patients”

Semantic MEDLINE Applications

➤ Discovery browsing



Acknowledgments

- Thomas C. Rindflesch, Ph.D.
- Michael J. Cairelli, D.O.
- Marcelo Fiszman, M.D., Ph.D.
- Graciela Rosemblat, Ph.D.
- Dongwook Shin, Ph.D.
- T. Elizabeth Workman, Ph.D., M.L.I.S.