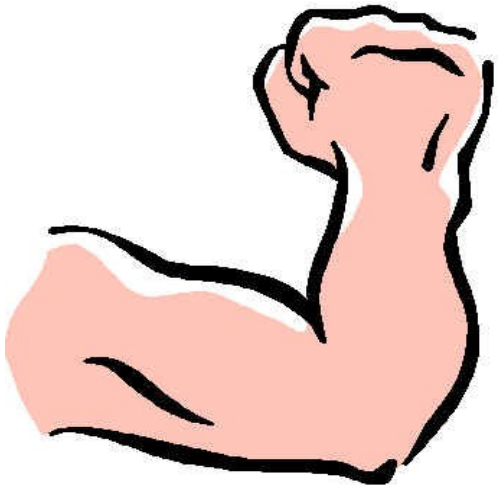
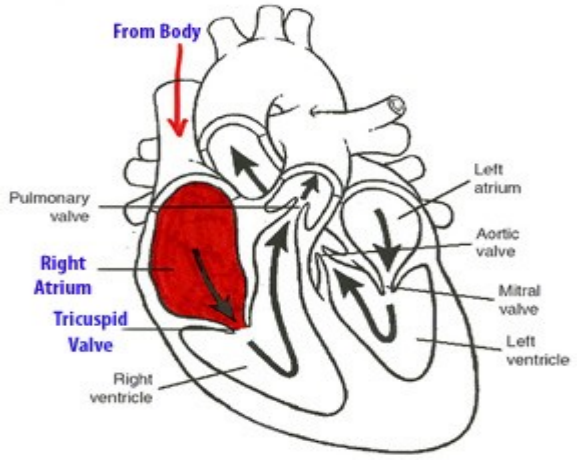


Right Arm versus Right Atrium:



How to distinguish between the two.

By Bridget T. McInnes
24 March 2011

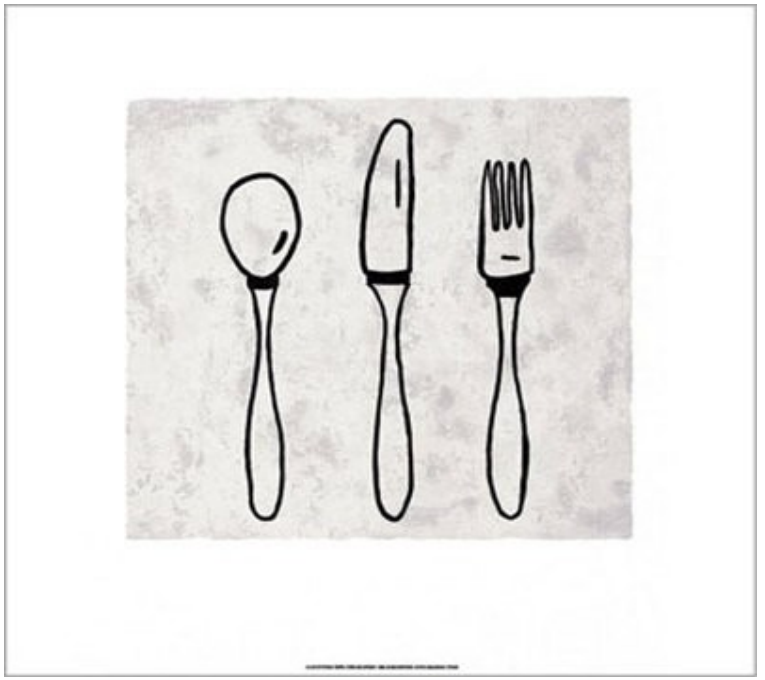
GOAL

- Present a method to disambiguate between terms and acronyms given the context in which they are used … using measures of semantic similarity and relatendess.

- Present:
  - A brief overview of semantic similarity and relatedness measures
  - Term and Acronym Disambiguation

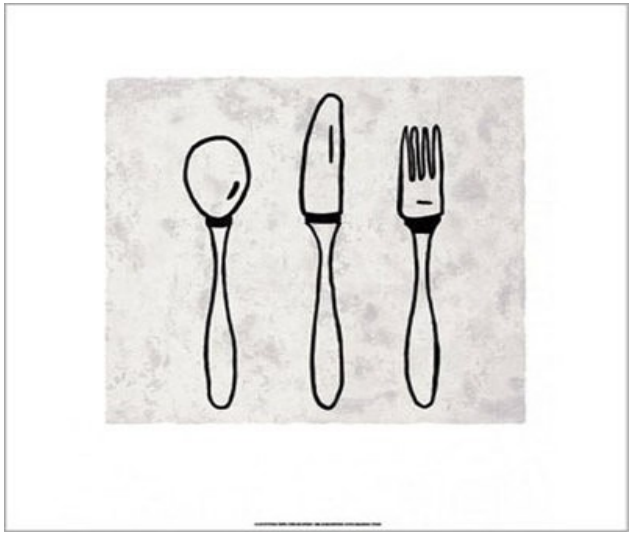
Similarity and Relatedness Measures

Measures the degree of similarity or relatedness between two terms/concepts.



Similarity Measures

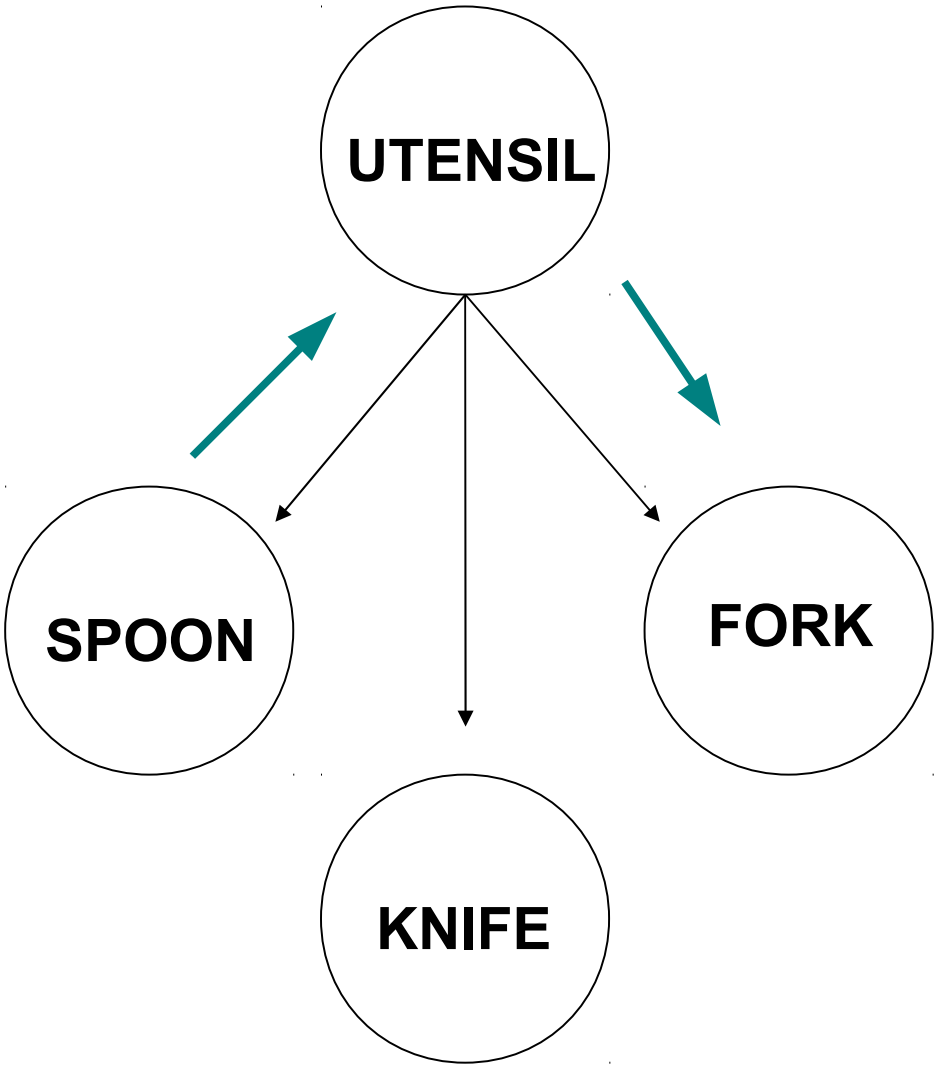
- Two types:
  - Path-based measures
  - Information content (IC)-based measures



Path-based Similarity Measures

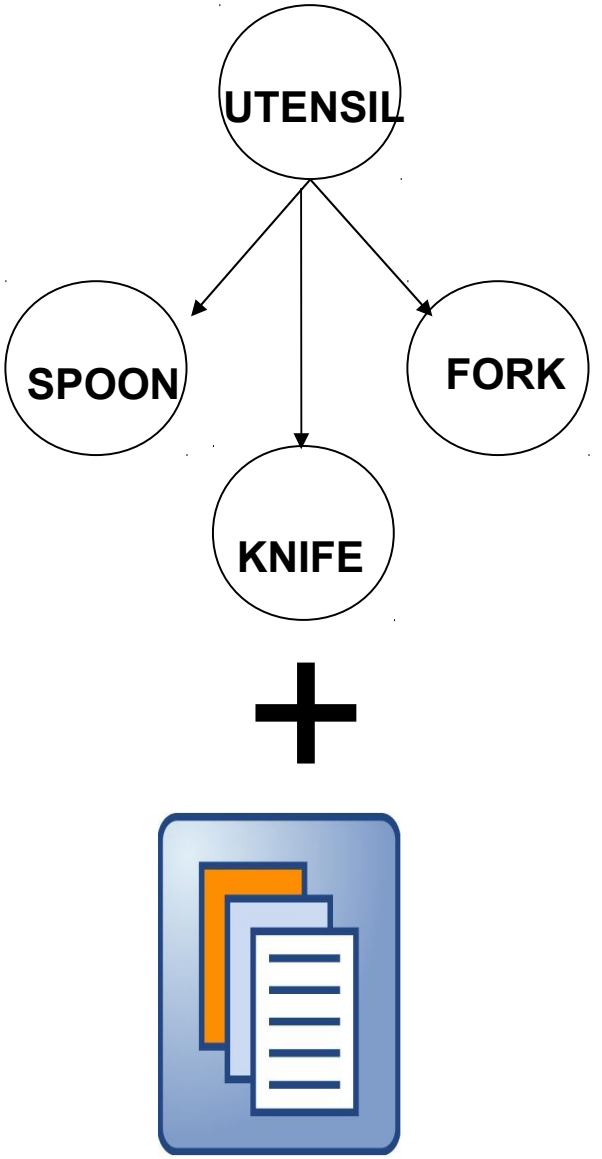
- Rely on the shortest path between two concepts in a taxonomy

- Path Measures:
  - Rada, et al. 1989
  - Leacock and Chodorow 1998
  - Wu and Palmer 1994
  - Nguyen and Al-Mubaid 2006



IC-based Similarity Measures

- Extended Path-based measure by incorporating the probability of the concept occurring in a text
- Information Content
  - Negative Log of the Probability of the Concept
- Measures Proposed by:
  - Resnik 1995
  - Jaing and Conrath 1997
  - Lin 1998



Relatedness Measures

- DO NOT rely on a taxonomy
- Rely on the terms and words that overlap in their definition

Needle: A **sewing needle** is a long slender tool with a pointed tip

Thread: A thread is a kind of yarn used with a **needle** for **sewing**

Measures: Lesk 1986

UMLS::Similarity

- Freely available open source Perl software package
- Used to determine the similarity/relatedness between biomedical and clinical concepts in the Unified Medical Language System
- Implements:
  - Path-based Similarity Measures
  - IC-based Similarity Measures
  - Relatedness Measure
- Note:
  - The path and definition information are obtained from the UMLS.

Unified Medical Language System

- Lexical database containing > 1.5 million clinical and biomedical concepts from over 100 terminology sources:
    - SNOMED CT
    - Medical Subject Headings (MSH)

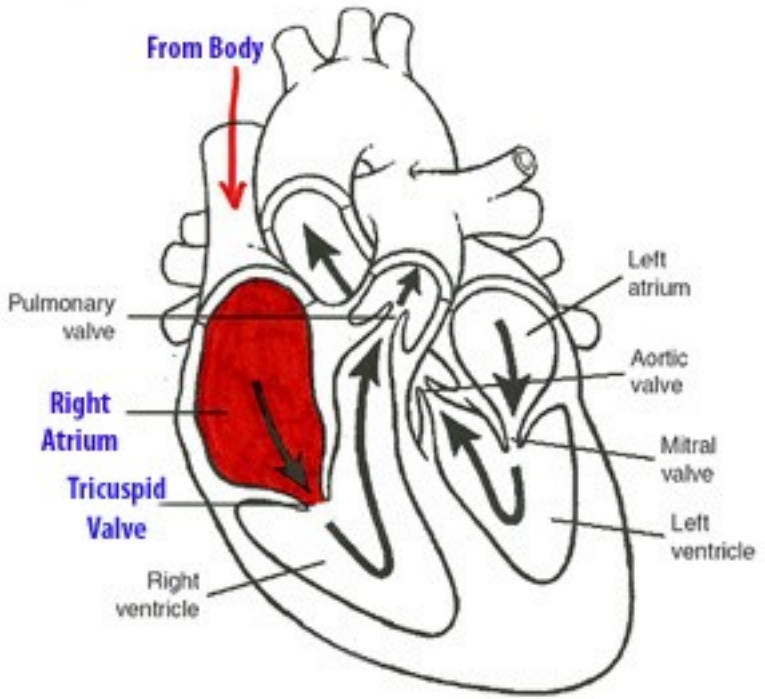
- UMLS::Similarity uses the path and definition information from the individual terminology sources as well as combinations sources in its implementation of the measures

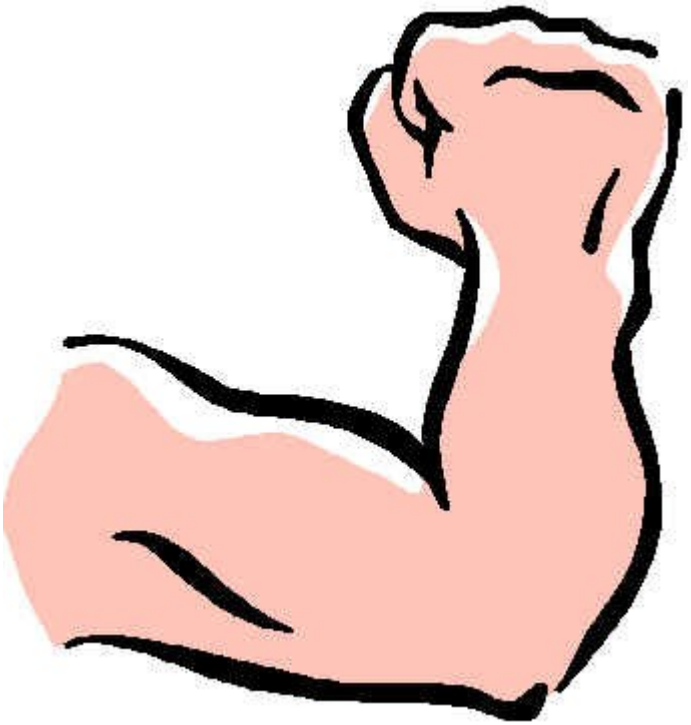
Using UMLS::Similarity measures for the purpose of term and acronym disambiguation

Term Disambiguation

The **bat** flew through the night

Acronym Disambiguation





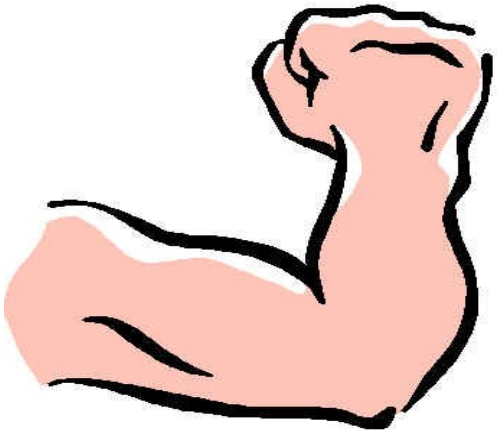
The **RA** incision is closed by
sewing it to the other edge of the left incision

UMLS::SenseRelate Algorithm

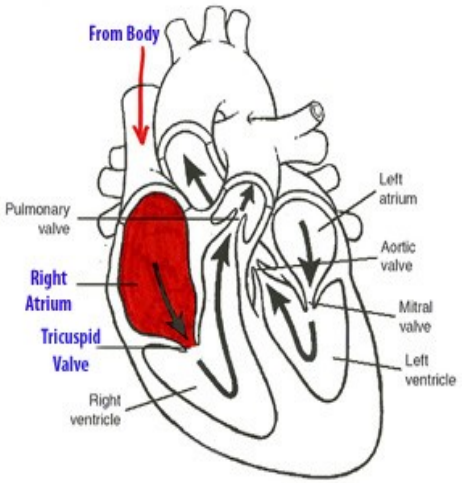
- Method:
  - Each possible sense/acronym is assigned a score by summing the similarity or relatedness between it and the terms surrounding the ambiguous word (or acronym)
  - Similarity or relatedness between the terms is obtained using UMLS::Similarity package
- Assumption:
  - The ambiguous word (or acronym) will be used in the sense that is most related to the senses of the terms that surrounds it

UMLS::SenseRelate Example

RIGHT ARM



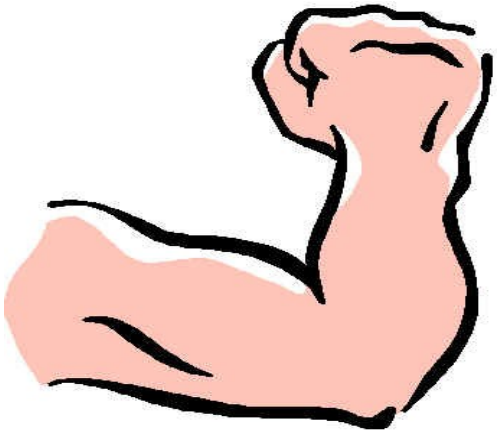
RIGHT ATRIUM



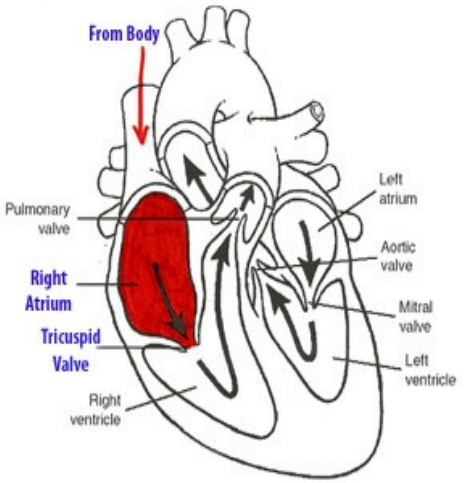
The **RA** incision is closed by
sewing it to the other edge of the left incision

UMLS::SenseRelate Example

RIGHT ARM

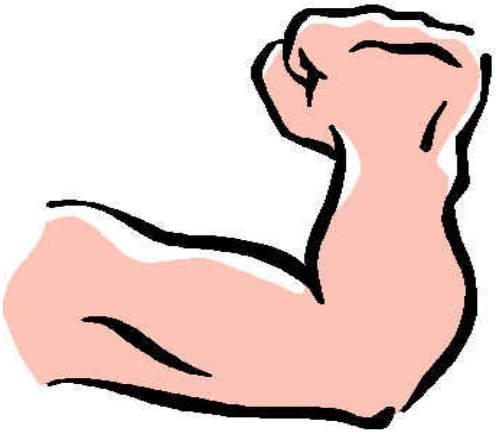


RIGHT ATRIUM



The **RA** *incision* is *closed* by *sewing* it to the other *edge* of the *left incision*

Sense: Right Arm



0.1 INCISION

0.2 CLOSE

0.3 SEWING

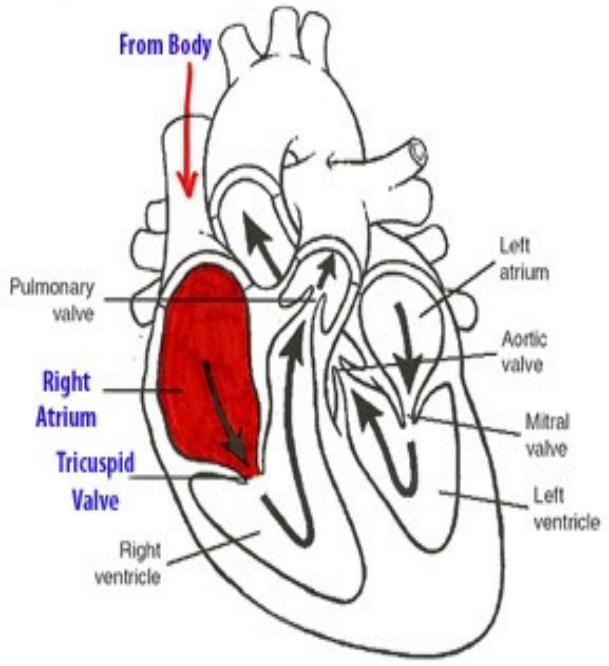
0.1 EDGE

0.5 LEFT INCISION

SCORE = 0.1 + 0.2 + 0.3 + 0.1 + 0.5

= 1.2

SENSE: Right Atrium



0.5 INCISION

0.2 CLOSE

0.2 SEWING

0.3 EDGE

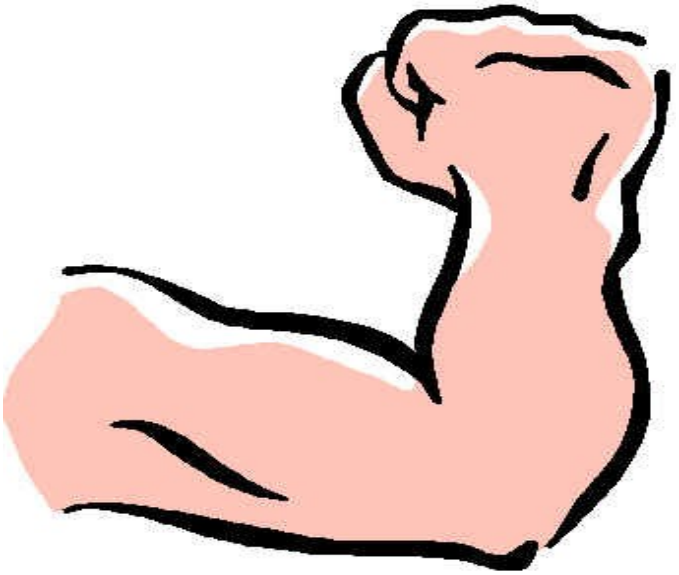
0.5 LEFT INCISION

SCORE = 0.5 + 0.2 + 0.2 + 0.3 + 0.5

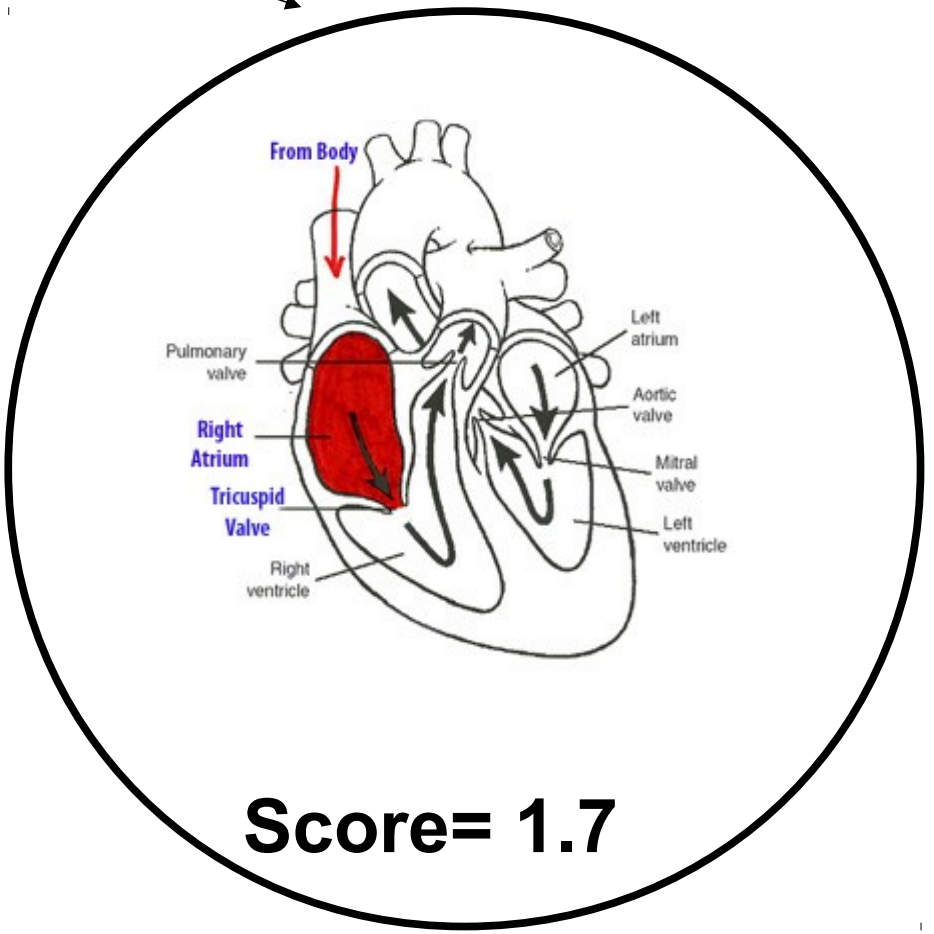
= 1.7

UMLS::SenseRelate Example

The **RA** incision is closed by
sewing to the other edge of the left incision



Score= 1.2



Score= 1.7

Mapping Details

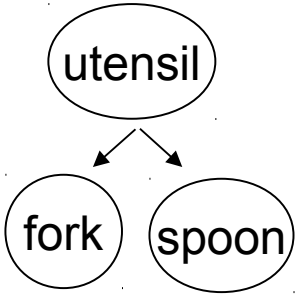
- Mapping surrounding terms to concepts
  - Terms are identified using the Specialist Lexicon
  - The mappings are done using a dictionary lookup in the Unified Medical Language System (UMLS)
    - You can already see the potential future work here ....

UMLS::SenseRelate Evaluation Data

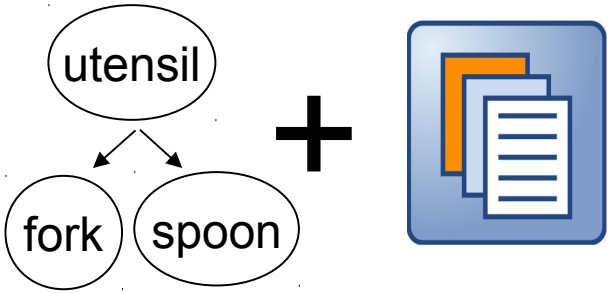
- MSH-WSD
  - Medline abstracts containing either an ambiguous term or acronym from the 2009 baseline
  - Each ambiguous term/acronym was assigned a concept from the Medical Subject Headings source (MSH) in the UMLS
  - Consists of:
    - Acronyms: 106
    - Terms: 88
    - Mixture of Terms/Acronyms: 9
      - cold:
        - temperature
        - Chronic Obstructive Airway Disease

Similarity Measures

- Path-based Similarity Measure
  - Taxonomy MSH in 2009AB UMLS



- IC-based Similarity Measure
  - Taxonomy: MSH in 2009AB UMLS
  - Probability Information: UMLSonMedline
    - Frequency counts for 2009AB concepts
    - 2009 Medline baseline

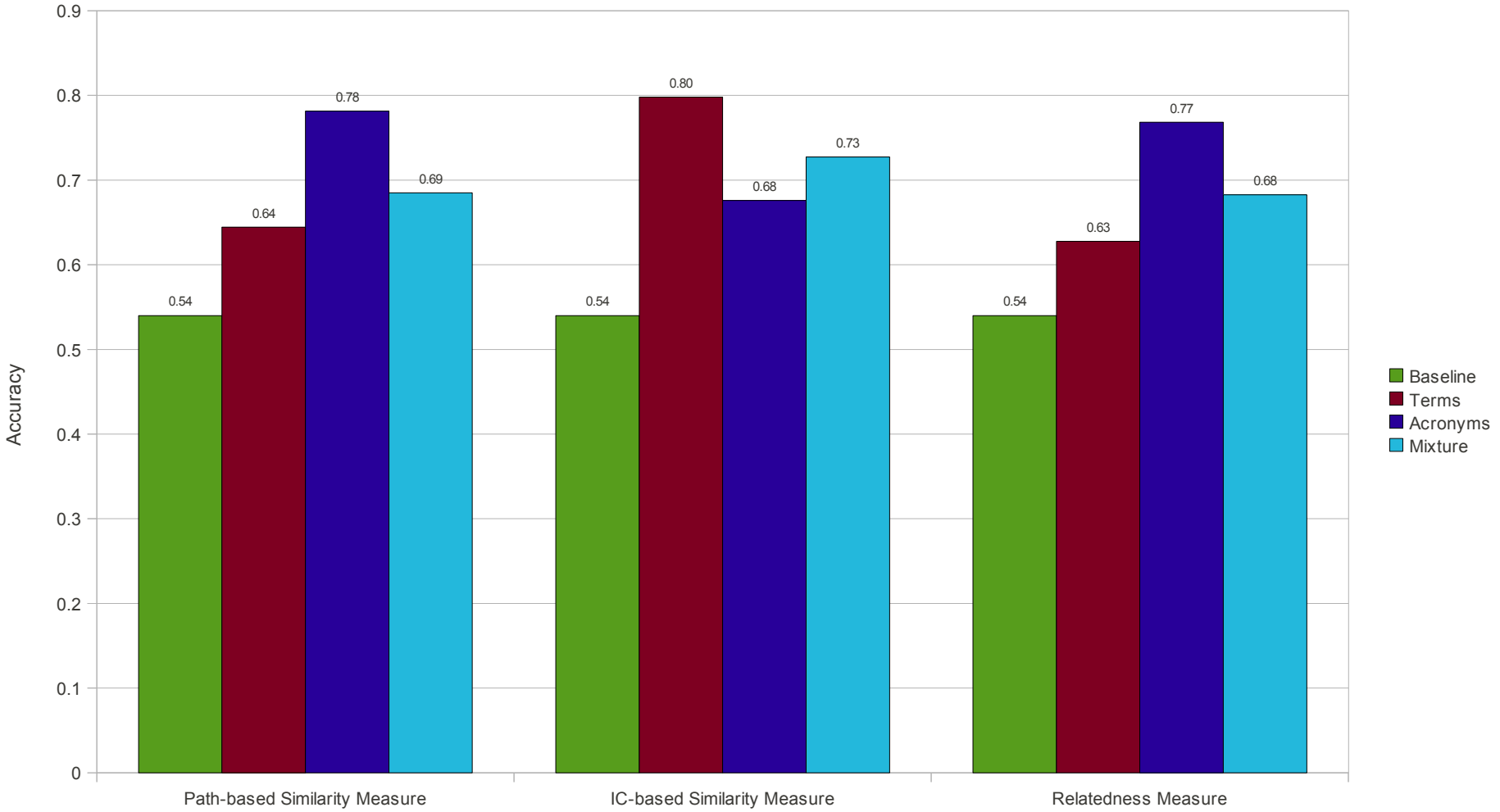


- Relatedness Measure
  - Definitions: 2009AB UMLS

NEEDLE: a long slender tool ...

THREAD: a kind of yarn ...

UMLS::SenseRelate Results



UMLS::SenseRelate Assumption

The ambiguous word or acronym will be used in the sense that is most related to the senses of the terms that surround it

UMLS::SenseRelate Error Analysis

- Assumption does not always hold true
  - Lawsonia
    - Bacteria Genus
    - Plant Genus
  - Assigned senses randomly indicating that the terms surrounding *lawsonia* are not more related to its correct sense .... but .....

UMLS::SenseRelate Conclusions

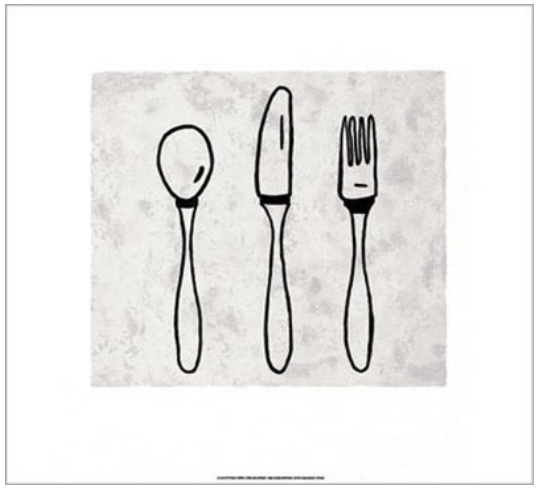
- In general, our assumption does hold true:
  - Ambiguous words or acronyms will be used in the sense the is most related to the senses that surrounds it

- **Conclude**: Semantic similarity and relatedness measures can be used for the purpose of disambiguation

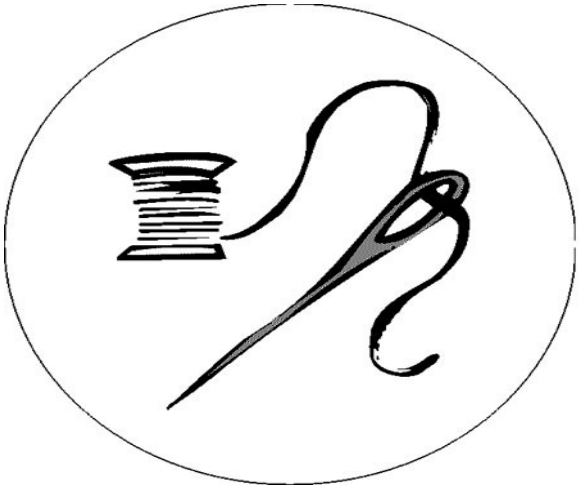
If you are going to remember anything from this talk ...

#1 UMLS::Similarity

Similarity Measures

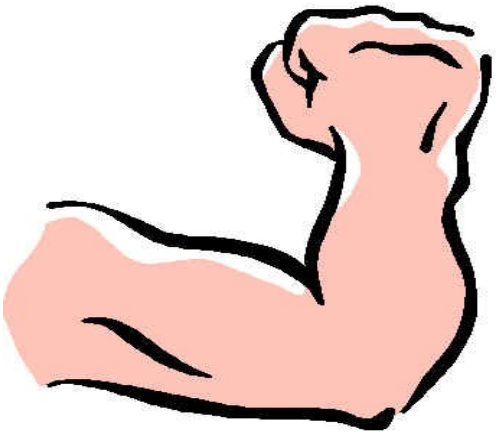


Relatedness Measures

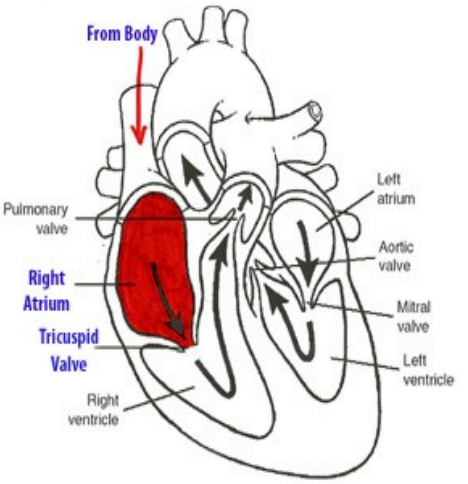


If you are going to remember anything from this talk ...

#1 UMLS::SenseRelate



RA?



Download

- UMLS::SenseRelate package

  http://search.cpan.org/dist/UMLS-SenseRelate

- UMLS::Similarity package

  http://search.cpan.org/dist/UMLS-Similarity

- UMLS::Similarity Web Interface

  http://atlas.ahc.umn.edu/cgi-bin/umls_similarity.cgi