

Measuring Semantic Relatedness Using a Medical Taxonomy

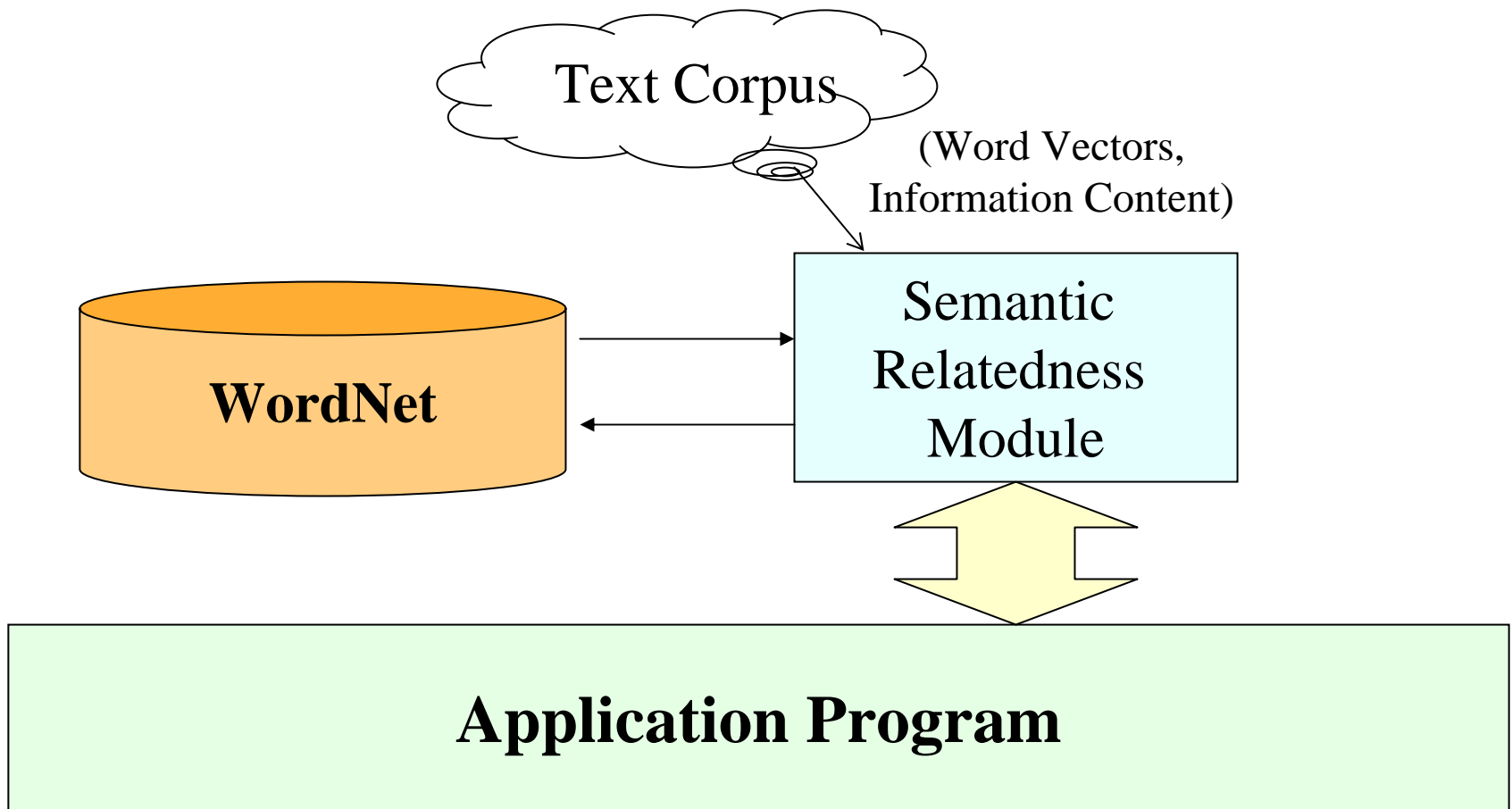
Siddharth Patwardhan

10/08/2003

WordNet::Similarity v0.05

- **Perl implementation of measures of semantic relatedness.**
- **Distributed on CPAN.**
- **WordNet-based measures.**
- **Some of the measures combine the structure and content of WordNet with statistical information from corpora.**

WordNet::Similarity – Implementation Details



Quick Recap of the Measures

- **Simple Edge Counts...**
 - Leacock Chodorow.
- **Information Content Based...**
 - Resnik
 - Lin
 - Jiang Conrath
- **Others...**
 - Gloss Overlaps
 - Vector

Using a Medical Taxonomy

- **A number of semantic networks, similar to WordNet, exist in the Medical Domain.**
- **By replacing WordNet with a semantic network like SNOMED CT or MeSH, we could measure the semantic relatedness of medical terms.**

Applications

- **Medical Informatics involves a great deal of text processing.**
- **Document Clustering.**
- **Retrieving relevant information.**
- **Ontology Construction?**

Resources – SNOMED

- **An ontology of clinical terms.**
- **Very similar to WordNet (concepts + relationships).**
- **Does not contain definitions.**
- **Predominant *is-a* hierarchy.**
- **Single root node.**

Resources – UMLS

- **An effort by NLM to provide a single complete repository medical taxonomies.**
- **Combine a large number of taxonomies (SNOMED, MeSH, ICD,...).**
- **Also, have manually created links between the nodes of different taxonomies.**

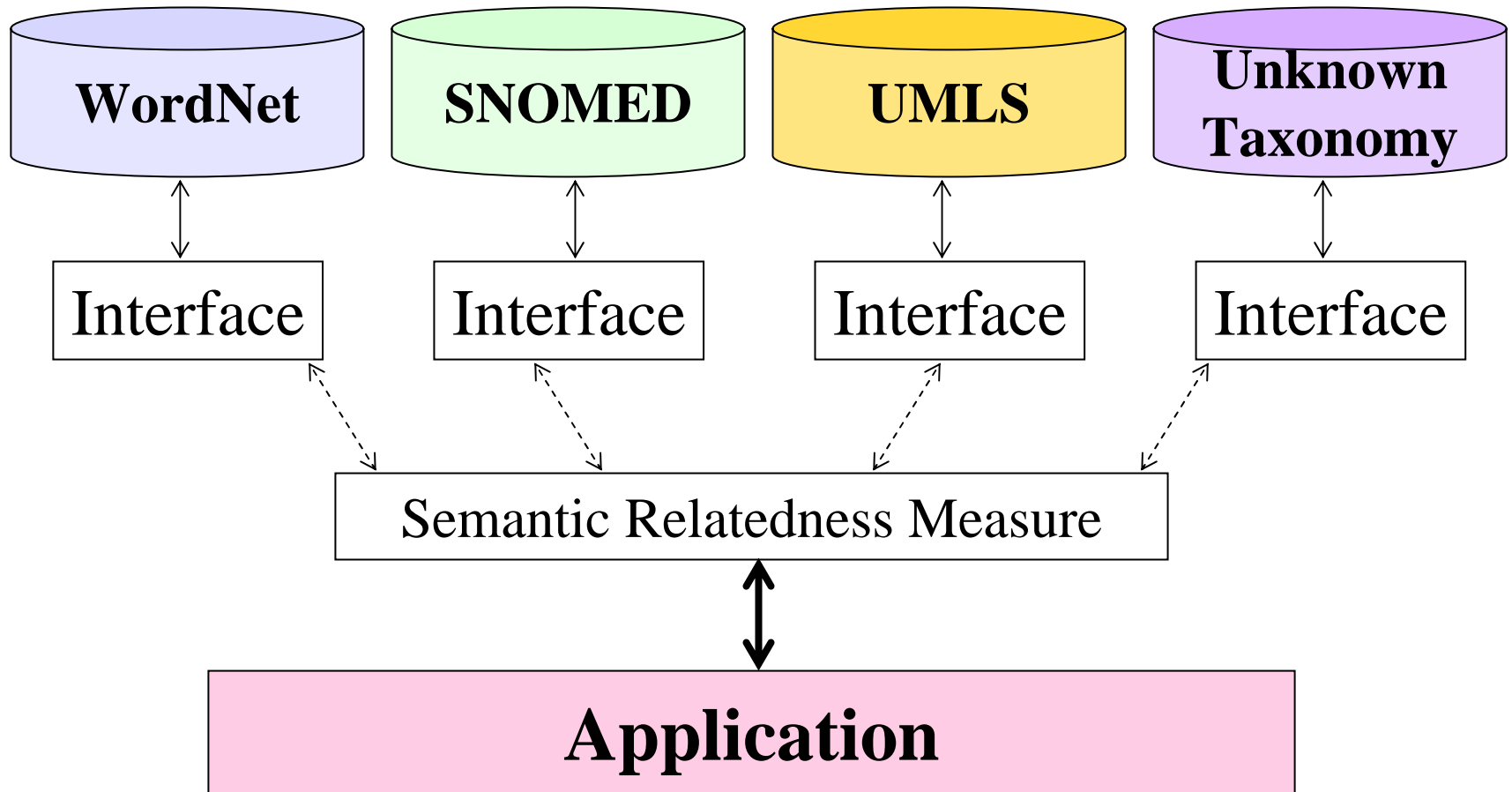
Resources – Patient Data

- **Hospitals like Mayo collect large amounts of patient data in the form of**
 - **Patient diagnoses.**
 - **Tests results.**
 - **Prescriptions, etc.**
- **It is plain text, since it is generated by physicians.**
- **Serves as a large corpus of medical text.**

A Generic Interface

- Instead of making the measures SNOMED specific or WordNet specific, we create a generic interface to the measures, such that any network may be plugged in, and be used for measuring relatedness.
- All the taxonomy specific details go into the interface.

Using a Medical Taxonomy



What is the interface?

The interface is basically a Perl object, which has a set of pre-decided methods.

```
new( )
taxonomy( CONCEPT )
exists( CONCEPT )
pathsToRoot( CONCEPT )
depth( TAXONOMY )
numberOfTaxonomies( )
root( TAXONOMY )
version( )
relations( )
query( CONCEPT, REL )
querySense( CONCEPT, REL )
queryWord( CONCEPT, REL )
getConcepts( TERM )
getConceptList( )
getTerms( CONCEPT )
getTermList( )
getError( )
```

Snomed::Interface

- **Created an interface to SNOMED.**
- **For term-to-concept mapping, in addition to using the terms in SNOMED, we used a database of clusters of terms mapped to concepts.**
- **These clusters were also used as “definitions” of concepts in some of the measures.**

Using the modules

```
#!/usr/bin/perl -w

use Snomed::Interface;
use WordNet::Similarity::vector;

$sm = Snomed::Interface->new();
$measure = WordNet::Similarity::vector->new($sm);

$rel = $measure->getRelatedness("118170006", "363708005");

($c1) = $sm->getConcepts("diabetes");
($c2) = $sm->getConcepts("blood sugar");

$rel = $measure->getRelatedness($c1, $c2);
```

Evaluation

- **This is a hard task, and we have not yet evaluated the usefulness of “ported” measures.**
- **We made an attempt at it.**
- **We had a physician create a list of medical term pairs.**
- **We had 14 experts in the medical domain score these pairs.**

Evaluation

- **It turned out that the experts did not agree much on the relatedness of the pairs.**
- **We haven't yet built an application that uses the measures – but this, in my mind, would be a better way to evaluate it potential in a real world task.**

Future Work

- **Main task to determine if the measures in the medical domain are any good at all.**
- **Using these in an application.**
- **Exploring other domains.**

Conclusions

- **We were able to generalize the semantic relatedness measures to the Medical Informatics field.**
- **We pulled apart the strong tie of the measures to WordNet.**