Ontology Alignment
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current issues
Ontologies in biomedical research

- many biomedical ontologies
e.g. GO, OBO, SNOMED-CT
- practical use of biomedical ontologies
e.g. databases annotated with GO

```
GENE ONTOLOGY (GO)

immune response
i- acute-phase response
i- anaphylaxis
i- antigen presentation
i- antigen processing
i- cellular defense response
i- cytokine metabolism
  i- cytokine biosynthesis
    synonym cytokine production
    ...
    p- regulation of cytokine biosynthesis
    ...
    ...
i- B-cell activation
i- B-cell differentiation
i- B-cell proliferation
i- cellular defense response
  ...
i- T-cell activation
  i- activation of natural killer cell activity
  ...
```
## Ontologies with overlapping information

### GENE ONTOLOGY (GO)
- immune response
  - acute-phase response
  - anaphylaxis
  - antigen presentation
  - antigen processing
  - cellular defense response
  - cytokine metabolism
    - cytokine biosynthesis
      - synonym cytokine production
      - regulation of cytokine biosynthesis
      -... 
    -... 
  - B-cell activation
    - B-cell differentiation
    - B-cell proliferation
  - cellular defense response
  -... 
- T-cell activation
  - activation of natural killer cell activity
  -... 

### SIGNAL-ONTOLOGY (SigO)
- Immune Response
  - Allergic Response
  - Antigen Processing and Presentation
  - B Cell Activation
  - B Cell Development
  - Complement Signaling
    - synonym complement activation
  - Cytokine Response
  - Immune Suppression
  - Inflammation
  - Intestinal Immunity
  - Leukotriene Response
    - Leukotriene Metabolism
    - Natural Killer Cell Response
  - T Cell Activation
  - T Cell Development
  - T Cell Selection in Thymus
Ontologies with overlapping information

- Use of multiple ontologies
  - custom-specific ontology + standard ontology
  - different views over same domain
  - overlapping domains

- Bottom-up creation of ontologies
  experts can focus on their domain of expertise

→ important to know the inter-ontology relationships
**GENE ONTOLOGY (GO)**

- immune response
  - acute-phase response
  - anaphylaxis
  - antigen presentation
  - antigen processing
  - cellular defense response
  - cytokine metabolism
    - cytokine biosynthesis
      synonym cytokine production
      ...
    - p-regulation of cytokine biosynthesis
      ...
    ...
  - B-cell activation
    - B-cell differentiation
    - B-cell proliferation
    - cellular defense response
      ...
  - T-cell activation
    - activation of natural killer cell activity
      ...

**SIGNAL-ONTOLOGY (SigO)**

- Immune Response
  - Allergic Response
  - Antigen Processing and Presentation
  - B Cell Activation
  - B Cell Development
  - Complement Signaling
    synonym complement activation
  - Cytokine Response
  - Immune Suppression
  - Inflammation
  - Intestinal Immunity
  - Leukotriene Response
    - Leukotriene Metabolism
  - Natural Killer Cell Response
  - T Cell Activation
  - T Cell Development
  - T Cell Selection in Thymus
Ontology Alignment

Defining the relations between the terms in different ontologies
Ontology Alignment

- Ontology alignment
- **Ontology alignment strategies**
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current issues
An Alignment Framework
Classification

- According to input
  - KR: OWL, UML, EER, XML, RDF, …
  - components: concepts, relations, instance, axioms

- According to process
  - What information is used and how?

- According to output
  - 1-1, m-n
  - Similarity vs explicit relations (equivalence, is-a)
  - confidence
Preprocessing
Preprocessing

For example,

- Selection of features
- Selection of search space
Matchers
Matcher Strategies

- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-based strategies
- Instance-based strategies
- Use of auxiliary information

GO: Complement Activation

SigO: complement signaling
    synonym complement activation
Example matchers

- **Edit distance**
  - Number of deletions, insertions, substitutions required to transform one string into another
  - `aaaa` → `baab`: edit distance 2

- **N-gram**
  - N-gram : N consecutive characters in a string
  - Similarity based on set comparison of n-grams
  - `aaaa` : `{aa, aa, aa}`; `baab` : `{ba, aa, ab}`
Matcher Strategies

- Strategies based on linguistic matching
- **Structure-based strategies**
- Constraint-based strategies
- Instance-based strategies
- Use of auxiliary information
Example matchers

- Propagation of similarity values
- Anchored matching
Example matchers

- Propagation of similarity values
- Anchored matching
Example matchers

- Propagation of similarity values
- Anchored matching
Matcher Strategies

- Strategies based on linguistic matching
- Structure-based strategies
- **Constraint-based approaches**
- Instance-based strategies
- Use of auxiliary information
Matcher Strategies

- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-based approaches
- Instance-based strategies
- Use of auxiliary information
Example matchers

- Similarities between data types
- Similarities based on cardinalities
Matcher Strategies

- Strategies based on linguistic matching
- Structure-based strategies
- Constraint-based approaches
- Instance-based strategies
- Use of auxiliary information

Diagram:
- Instance corpus
- Ontology
Example matchers

- Instance-based
- Use life science literature as instances

- Structure-based extensions
Learning matchers – instance-based strategies

- **Basic intuition**
  
  A similarity measure between concepts can be computed based on the probability that documents about one concept are also about the other concept and vice versa.

- **Intuition for structure-based extensions**
  
  Documents about a concept are also about their super-concepts.
  
  (No requirement for previous alignment results.)
Learning matchers - steps

- Generate corpora
  - Use concept as query term in PubMed
  - Retrieve most recent PubMed abstracts
- Generate text classifiers
  - One classifier per ontology / One classifier per concept
- Classification
  - Abstracts related to one ontology are classified by the other ontology’s classifier(s) and vice versa
- Calculate similarities
Basic Naïve Bayes matcher

- Generate corpora
- Generate classifiers
  - Naive Bayes classifiers, one per ontology
- Classification
  - Abstracts related to one ontology are classified to the concept in the other ontology with highest posterior probability $P(C|d)$
- Calculate similarities

$$\text{sim}(C_1, C_2) = \frac{nW \cdot \text{bc}(C_1, C_2) + nW \cdot \text{bc}(C_2, C_1)}{nD(C_1) + nD(C_2)}$$
Basic Support Vector Machines matcher

- Generate corpora
- Generate classifiers
  - SVM-based classifiers, one per concept
- Classification
  - Single classification variant: Abstracts related to concepts in one ontology are classified to the concept in the other ontology for which its classifier gives the abstract the highest positive value.
  - Multiple classification variant: Abstracts related to concepts in one ontology are classified all the concepts in the other ontology whose classifiers give the abstract a positive value.
- Calculate similarities

\[
\frac{n_{SV_{MC-C_2}}(C_1, C_2) + n_{SV_{MC-C_1}}(C_2, C_1)}{n_D(C_1) + n_D(C_2)}
\]
Structural extension ‘Cl’

- Generate classifiers
  - Take (is-a) structure of the ontologies into account when building the classifiers
  - Extend the set of abstracts associated to a concept by adding the abstracts related to the sub-concepts
Structural extension ‘Sim’

- Calculate similarities
  - Take structure of the ontologies into account when calculating similarities
  - Similarity is computed based on the classifiers applied to the concepts and their sub-concepts

\[
Sim_{\text{struct}}(C_1, C_2) = \frac{\sum_{C_3 \subseteq C_1} \text{sim}(C_3, C_2) + \sum_{C_3 \subseteq C_2} \text{sim}(C_3, C_1)}{\sum_{C_3 \subseteq C_1} \#(C_3) + \sum_{C_3 \subseteq C_2} \#(C_3)}
\]
Matcher Strategies

- Strategies based linguistic
- Structure-based strategies
- Constraint-based approaches
- Instance-based strategies
- Use of auxiliary information
Example matchers

- Use of WordNet
  - Use WordNet to find synonyms
  - Use WordNet to find ancestors and descendants in the is-a hierarchy

- Use of Unified Medical Language System (UMLS)
  - Includes many ontologies
  - Includes many alignments (not complete)
  - Use UMLS alignments in the computation of the similarity values
<table>
<thead>
<tr>
<th>System</th>
<th>linguistic</th>
<th>structure</th>
<th>constraints</th>
<th>instances</th>
<th>auxiliary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArtGen</td>
<td>name</td>
<td>parents, children</td>
<td>domain specific documents</td>
<td></td>
<td>WordNet</td>
</tr>
<tr>
<td>ASCO</td>
<td>name, label description</td>
<td>parents, children, siblings, path from root</td>
<td></td>
<td></td>
<td>WordNet</td>
</tr>
<tr>
<td>Chimaera</td>
<td>name</td>
<td>parents, children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCA-Merge</td>
<td>name</td>
<td></td>
<td>domain specific documents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOAM</td>
<td>name, label</td>
<td>parents, children</td>
<td>equivalence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUE</td>
<td>name</td>
<td>neighborhood</td>
<td>instances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCONET</td>
<td>name</td>
<td>parents, children</td>
<td></td>
<td></td>
<td>WordNet</td>
</tr>
<tr>
<td>IF-Map</td>
<td></td>
<td></td>
<td>instances</td>
<td>a reference ontology</td>
<td></td>
</tr>
<tr>
<td>iMapper</td>
<td></td>
<td>leaf, non-leaf, children, related node</td>
<td>domain, range</td>
<td>instances</td>
<td>WordNet</td>
</tr>
<tr>
<td>OntoMapper</td>
<td></td>
<td>parents, children</td>
<td></td>
<td>documents</td>
<td></td>
</tr>
<tr>
<td>(Anchor-) PROMPT</td>
<td>name</td>
<td>direct graphs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAMBO</td>
<td>name, synonym</td>
<td>is-a and part-of, descendants and ancestors</td>
<td>domain specific documents</td>
<td></td>
<td>WordNet, UMLS</td>
</tr>
<tr>
<td>S-Match</td>
<td>label</td>
<td>path from root</td>
<td>semantic relations codified in labels</td>
<td></td>
<td>WordNet</td>
</tr>
</tbody>
</table>
Combinations
Combination Strategies

- Usually weighted sum of similarity values of different matchers
- Maximum of similarity values of different matchers
Filtering techniques

- Threshold filtering

Pairs of concepts with similarity higher or equal than threshold are alignment suggestions

```
( 2, B )
( 3, F )
( 6, D )
( 4, C )
( 5, C )
( 5, E )
```

......
Filtering techniques

- Double threshold filtering
  
  (1) Pairs of concepts with similarity higher than or equal to upper threshold are alignment suggestions.
  
  (2) Pairs of concepts with similarity between lower and upper thresholds are alignment suggestions if they make sense with respect to the structure of the ontologies and the suggestions according to (1).

```
(2, B)
(3, F)
(6, D)
```

```
upper-th (4, C)
(5, C)
```

```
lower-th (5, E)
......
```

Example alignment system

SAMBO – matchers, combination, filter

<table>
<thead>
<tr>
<th>start</th>
<th>relation</th>
<th>concept</th>
<th>finish</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Align Concept in Ontology-1 and Ontology-2

Matchers:
- Terminology
- Term. + WordNet
- Domain (UMLS)
- Learning
- Structure

Threshold: 0.6

Start

Finish
Example alignment system
SAMBO – suggestion mode
Example alignment system
SAMBO – manual mode
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current issues
Evaluation measures

- Precision:
  \[
  \frac{\text{# correct suggested alignments}}{\text{# suggested alignments}}
  \]

- Recall:
  \[
  \frac{\text{# correct suggested alignments}}{\text{# correct alignments}}
  \]

- F-measure: combination of precision and recall
Ontology Alignment Evaluation Initiative
OAEI

- Since 2004
- Evaluation of systems
- Different tracks
  - comparison: benchmark (open)
  - expressive: anatomy (blind), fisheries (expert)
  - directories and thesauri: directory, library, crosslingual resources (blind)
  - consensus: conference
OAEI

- Evaluation measures
  - Precision/recall/f-measure
  - recall of non-trivial mappings
  - full / partial golden standard
OAEI 2007

- 17 systems participated
  - benchmark (13)
    - ASMOV: $p = 0.95, r = 0.90$
  - anatomy (11)
    - AOAS: $f = 0.86, r^+ = 0.50$
    - SAMBO: $f = 0.81, r^+ = 0.58$
  - library (3)
    - Thesaurus merging: FALCON: $p = 0.97, r = 0.87$
    - Annotation scenario:
      - FALCON: $pb = 0.65, rb = 0.49, pa = 0.52, ra = 0.36, Ja = 0.30$
      - Silas: $pb = 0.66, rb = 0.47, pa = 0.53, ra = 0.35, Ja = 0.29$
  - directory (9), food (6), environment (2), conference (6)
OAEI 2008 – anatomy track

Mouse anatomy: 2744 terms
NCI-anatomy: 3304 terms
Mappings: 1544 (of which 934 ‘trivial’)

Tasks
1. Align and optimize f
2-3. Align and optimize p / r
4. Align when partial reference alignment is given and optimize f
OAEI 2008 – anatomy track#1

- 9 systems participated
- SAMBO
  - p=0.869, r=0.836, r+=0.586, f=0.852
- SAMBOdtf
  - p=0.831, r=0.833, r+=0.579, f=0.832
- Use of TermWN and UMLS
OAEI 2008 – anatomy track#1

Is background knowledge (BK) needed?

Of the non-trivial mappings:

- Ca 50% found by systems using BK and systems not using BK
- Ca 13% found only by systems using BK
- Ca 13% found only by systems not using BK
- Ca 25% not found

Processing time:

hours with BK, minutes without BK
OAEI 2008 – anatomy track#4

Can we use given mappings when computing suggestions?

→ partial reference alignment given with all trivial and 50 non-trivial mappings

- SAMBO
  - $p=0.636 \rightarrow 0.660$, $r=0.626 \rightarrow 0.624$, $f=0.631 \rightarrow 0.642$

- SAMBO$dtf$
  - $p=0.563 \rightarrow 0.603$, $r=0.622 \rightarrow 0.630$, $f=0.591 \rightarrow 0.616$

(measures computed on non-given part of the reference alignment)
OAEI 2007-2008

- Systems can use only one combination of strategies per task
  - systems use similar strategies
    - text: string matching, tf-idf
    - structure: propagation of similarity to ancestors and/or descendants
    - thesaurus (WordNet)
    - domain knowledge important for anatomy task?
Evaluation of algorithms
Cases

□ GO vs. SigO

GO: 70 terms

<table>
<thead>
<tr>
<th>GO-immune defense</th>
<th>SigO-immune defense</th>
</tr>
</thead>
</table>

GO: 60 terms

<table>
<thead>
<tr>
<th>GO-behavior</th>
<th>SigO-behavior</th>
</tr>
</thead>
</table>

□ MA vs. MeSH

MA: 15 terms

<table>
<thead>
<tr>
<th>MA-nose</th>
<th>MeSH-nose</th>
</tr>
</thead>
</table>

MA: 77 terms

<table>
<thead>
<tr>
<th>MA-ear</th>
<th>MeSH-ear</th>
</tr>
</thead>
</table>

MA: 112 terms

<table>
<thead>
<tr>
<th>MA-eye</th>
<th>MeSH-eye</th>
</tr>
</thead>
</table>

MA: 45 terms

<table>
<thead>
<tr>
<th>MeSH-eye</th>
</tr>
</thead>
</table>
Evaluation of matchers

- Matchers
  
  Term, TermWN, Dom, Learn (Learn+structure), Struc

- Parameters
  
  Quality of suggestions: precision/recall
  Threshold filtering : 0.4, 0.5, 0.6, 0.7, 0.8
  Weights for combination: 1.0/1.2

KitAMO
(http://www.ida.liu.se/labs/iislab/projects/KitAMO)
Results

- Terminological matchers
Results

- Basic learning matcher (Naïve Bayes)

Naive Bayes slightly better recall, but slightly worse precision than SVM-single
SVM-multiple (much) better recall, but worse precision than SVM-single
Results

- Domain matcher (using UMLS)
Results

- Comparison of the matchers

  \[ CS_{TermWN} \supseteq CS_{Dom} \supseteq CS_{Learn} \]

- Combinations of the different matchers
  
  - combinations give often better results
  
  - no significant difference on the quality of suggestions for different weight assignments in the combinations
  (but: did not check yet for large variations for the weights)

- Structural matcher did not find (many) new correct alignments
  (but: good results for systems biology schemas SBML – PSI MI)
Evaluation of filtering

- Matcher
  TermWN

- Parameters
  Quality of suggestions: precision/recall
  Double threshold filtering using structure:
    Upper threshold: 0.8
    Lower threshold: 0.4, 0.5, 0.6, 0.7, 0.8
The precision for double threshold filtering with upper threshold 0.8 and lower threshold T is higher than for threshold filtering with threshold T.
Results

- The recall for double threshold filtering with upper threshold 0.8 and lower threshold $T$ is about the same as for threshold filtering with threshold $T$. 
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current issues
Recommending strategies - 1

- Use knowledge about previous use of alignment strategies
  - gather knowledge about input, output, use, performance, cost via questionnaires
  - Not so much knowledge available
  - OAEI

(Mochol, Jentzsch, Euzenat 2006)
Recommending strategies - 2

- Optimize
  - Parameters for ontologies, similarity assessment, matchers, combinations and filters
  - Run general alignment algorithm
  - User validates the alignment result
  - Optimize parameters based on validation

(Ehrig, Staab, Sure 2005)
Recommending strategies - 2

Tests

- travel in russia
  QOM: $r=0.618$, $p=0.596$, $f=0.607$
  Decision tree 150: $r=0.723$, $p=0.591$, $f=0.650$

- bibster
  QOM: $r=0.279$, $p=0.397$, $f=0.328$
  Decision tree 150: $r=0.630$, $p=0.375$, $f=0.470$

Decision trees better than Neural Nets and Support Vector Machines.
Recommending strategies - 3

- Based on inherent knowledge
  - Use the actual ontologies to align to find good candidate alignment strategies
  - User/oracle with minimal alignment work
  - Complementary to the other approaches

(Tan, Lambrix 2007)
Idea

- Select small segments of the ontologies
- Generate alignments for the segments (expert/oracle)
- Use and evaluate available alignment algorithms on the segments
- Recommend alignment algorithm based on evaluation on the segments
Framework

Segment Pair Selection Algorithm

Alignment Generator

Alignment Toolbox

evaluation result report

Recommended Alignment Strategy

Alignment Strategy Recommendation Algorithm

alignment strategies
matchers filters
combination algorithms
Experiment case
- Ontologies

- NCI thesaurus
  - National Cancer Institute, Center for Bioinformatics
  - Anatomy: 3495 terms

- MeSH
  - National Library of Medicine
  - Anatomy: 1391 terms
Experiment case - Oracle

- UMLS
  - Library of Medicine
  - Metathesaurus contains > 100 vocabularies
  - NCI thesaurus and MeSH included in UMLS
  - Used as approximation for expert knowledge
  - 919 expected mappings according to UMLS
Experiment case
– alignment strategies

- Matchers and combinations
  - N-gram (NG)
  - Edit Distance (ED)
  - Word List + stemming (WL)
  - Word List + stemming + WordNet (WN)
  - NG+ED+WL, weights 1/3 (C1)
  - NG+ED+WN, weights 1/3 (C2)

- Threshold filter
  - thresholds 0.4, 0.5, 0.6, 0.7, 0.8
Segment pair selection algorithms

- SubG
  - Candidate segment pair = sub-graphs according to is-a/part-of with roots with same name; between 1 and 60 terms in segment
  - Segment pairs randomly chosen from candidate segment pairs such that segment pairs are disjoint
Segment pair selection algorithms

- Clust - Cluster terms in ontology
  - Candidate segment pair is pair of clusters containing terms with the same name; at least 5 terms in clusters
  - Segment pairs randomly chosen from candidate segment pairs
Segment pair selection algorithms

- For each trial, 3 segment pair sets with 5 segment pairs were generated

- SubG: A1, A2, A3
  - 2 to 34 terms in segment
  - level of is-a/part-of ranges from 2 to 6
  - max expected alignments in segment pair is 23

- Clust: B1, B2, B3
  - 5 to 14 terms in segment
  - level of is-a/part-of is 2 or 3
  - max expected alignments in segment pair is 4
Segment pair alignment generator

- Used UMLS as oracle

Alignment toolbox

- Used KitAMO as toolbox
- Generates reports on similarity values produced by different matchers, execution times, number of correct, wrong, redundant suggestions
Recommendation algorithm

- Recommendation scores: F, F+E, 10F+E
  - F: quality of the alignment suggestions
    - average f-measure value for the segment pairs
  - E: average execution time over segment pairs, normalized with respect to number of term pairs
- Algorithm gives ranking of alignment strategies based on recommendation scores on segment pairs
Expected recommendations for F

- Best strategies for the whole ontologies and measure F:
  1. (WL, 0.8)
  2. (C1, 0.8)
  3. (C2, 0.8)
Results

SubG, F, SPS A1
Results

- Top 3 strategies for SubG and measure F:
  A1: 1. (WL, 0.8) (WL, 0.7) (C1, 0.8) (C2, 0.8)
  A2: 1. (WL, 0.8) 2. (WL, 0.7) 3. (WN, 0.7)
  A3: 1. (WL, 0.8) (WL, 0.7) (C1, 0.8) (C2, 0.8)

- Best strategy always recommended first
- Top 3 strategies often recommended
- (WL, 0.7) has rank 4 for whole ontologies
Results

- Top 3 strategies for Clust and measure F:
  B1: 1. (C2,0.7) 2. (ED,0.6) 3. (C2,0.6)
  B2: 1. (WL,0.8) (WL, 0.7) (C1,0.8) (C2,0.8)
  B3: 1. (C1,0.8) (ED,0.7) 3. (C1,0.7) (C2,0.7) (WL,0.7) (WN,0.7)

- Top strategies often recommended, but not always
  (WL,0.7) (C1,0.7) (C2,0.7) ranked 4,5,6 for whole ontologies
Results

- SubG gives better results than Clust
- Results improve when number of segments is increased
- 10F+E similar results as F
- F+E
  - WordNet gives lower ranking
  - Runtime environment has influence
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current issues
Ontology Alignment Framework
Partial Reference Alignment

- New setting for ontology alignment:
  - Portals with mappings
  - Iterative ontology alignment
  - Anatomy track, task 4 in OAEI 2008
    → In all these cases some correct mappings between terms in different ontologies are given or have been obtained.

- A partial reference alignment (PRA) is a subset of all correct mappings.
Partial Reference Alignment

- Research Problem:
  Can we use PRAs to obtain higher quality mapping suggestions in ontology alignment?
Partial Reference Alignment

- **Research Problem:**
  Can we use PRAs in the different parts of the framework to obtain higher quality mapping suggestions in ontology alignment?
Test cases

<table>
<thead>
<tr>
<th>DataSet</th>
<th>Concepts in Ontology 1</th>
<th>Concepts in Ontology 2</th>
<th>Mappings in RA</th>
<th>Mappings in PRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
<td>57</td>
<td>10</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Defense</td>
<td>69</td>
<td>17</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Nose</td>
<td>18</td>
<td>15</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Ear</td>
<td>78</td>
<td>39</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>Eye</td>
<td>113</td>
<td>45</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>Anatomy</td>
<td>2743</td>
<td>3304</td>
<td>1523</td>
<td>988</td>
</tr>
</tbody>
</table>

- Behavior, Defense: Gene Ontology – Signal Ontology
- Nose, Ear, Eye: Adult Mouse Anatomy - MeSH
- Anatomy: Adult Mouse Anatomy – NCI anatomy
Evaluation

- **Precision**: number of correct suggestions divided by number of suggestions
- **Recall**: number of correct suggestions divided by number of correct mappings
- **Recall-PRA**: number of correct suggestions not in PRA divided by number of correct mappings not in PRA
- **F-measure**: harmonic mean of precision and recall
## Algorithms

### Table 1. Alignment strategies

<table>
<thead>
<tr>
<th></th>
<th>preprocessing</th>
<th>matchers</th>
<th>combination</th>
<th>filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMBO</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>single threshold</td>
</tr>
<tr>
<td>SAMBO dtf</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>double threshold</td>
</tr>
<tr>
<td>mgPRA</td>
<td>partitioning</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>single threshold filter with PRA</td>
</tr>
<tr>
<td>mgfPRA</td>
<td>fixing and partitioning</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>single threshold filter with PRA</td>
</tr>
<tr>
<td>pmPRA</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>single threshold filter with PRA</td>
</tr>
<tr>
<td>fPRA</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>single threshold filter with PRA</td>
</tr>
<tr>
<td>dtfPRA</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>double threshold filter with PRA</td>
</tr>
<tr>
<td>pfPRA</td>
<td>none</td>
<td>TermWN + UMLSKSearch</td>
<td>maximum</td>
<td>filter based on EM and PRA filter with PRA</td>
</tr>
</tbody>
</table>
1. Use of PRA in the preprocessing step
Use of PRA in the preprocessing step

- Intuition
  During the preprocessing step, use mappings in PRA to partition the ontologies into mappable groups.

- Methods
  - mgPRA
  - mgfPRA
Use of PRA in the preprocessing step

- **mgPRA** (Mappable Groups with PRA)
  - **Strategy**
    - Find consistent group in PRA
    - Partition ontologies into mappable groups before aligning
  - **Example:**
Use of PRA in the preprocessing step

Partition Results

Consistent Group in PRA

(6, D)
(2, B)
(3, F)
Use of PRA in the preprocessing step

- mgfPRA (Mappable Groups and Fixing with PRA)

  - Strategy
    - ‘Fix’ the missing structural relationships, making the whole PRA a consistent group
    - Then, partition ontologies into mappable groups

  - Example:

```
Fixed Ontology 1

Ontology 1

Ontology 2
```
Use of PRA in the preprocessing step

Partition Results

Omology O

Fixed Omology F

(6, F)
(3, F)
(2, B)

PR4
Use of PRA in the preprocessing step

<table>
<thead>
<tr>
<th>Case</th>
<th>RA</th>
<th>PRA</th>
<th>Th</th>
<th>SAMBO</th>
<th>mgPRA</th>
<th>mgPRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>0.66/1/0.871</td>
<td>0.66/1/0.871</td>
<td>1/0.5/0.66/0</td>
</tr>
<tr>
<td>ID</td>
<td>4</td>
<td>4</td>
<td>0.4</td>
<td>0.5/0.75/0.6/0.75</td>
<td>0.4/0.62/0.5/0.25</td>
<td>0.41/0.62/0.5/0.25</td>
</tr>
<tr>
<td>face</td>
<td>4</td>
<td>4</td>
<td>0.4</td>
<td>0.75/0.75/0.75/0.75</td>
<td>1/0.62/0.75/0.25</td>
<td>1/0.62/0.75/0.25</td>
</tr>
<tr>
<td>cur</td>
<td>27</td>
<td>14</td>
<td>0.4</td>
<td>0.86/0.86/0.91/0.91</td>
<td>0.86/0.86/0.87/0.76</td>
<td>1/0.66/0.86/0.82</td>
</tr>
<tr>
<td>eye</td>
<td>27</td>
<td>15</td>
<td>0.4</td>
<td>0.86/0.86/0.86/0.86</td>
<td>0.86/0.86/0.86/0.86</td>
<td>1/0.48/0.86/0.86</td>
</tr>
<tr>
<td>Anatomy</td>
<td>1523</td>
<td>968</td>
<td>0.4</td>
<td>0.86/0.86/0.86/0.86</td>
<td>0.86/0.86/0.86/0.86</td>
<td>0.86/0.86/0.86/0.86</td>
</tr>
</tbody>
</table>

Table 3. Using the PRA in the preprocessing phase (precision/recall/f-measure/recall.)
Use of PRA in the preprocessing step

Result Analysis

- For threshold 0.4, there are no conclusive results.
- For thresholds 0.6 and 0.8,
  - mgPRA and mgfPRA almost always have equal or higher precision than SAMBO.
  - mgPRA almost always has equal or higher recall than SAMBO.
  - mgfPRA almost always has equal or lower recall than SAMBO and mgPRA.
Use of PRA in the preprocessing step

- Why does mgfPRA perform worse than mgPRA?

Incorrect use of the structural relation.

For instance, in dataset nose, one source ontology uses the structural relation to define both is-a and part-of.

‘Fixing’ the ontology may therefore be wrong.

For instance, the mapping (nose, nose) may lead to introducing is-a relations between nose and its parts.
2. Use of PRA in the matcher
Use of PRA in a matcher

Observation

Some correct mappings share a similar linguistic pattern.

Examples from PRA of Anatomy

- (lumbar vertebra 5, l5 vertebra) and (thoracic vertebra 11, t11 vertebra)
- (forebrain, fore brain) and (gallbladder, gall bladder)
- (stomach body, body stomach) and (stomach fundus, fundus stomach)

Linguistic similarity vectors for (lumbar vertebra 5, l5 vertebra) and (thoracic vertebra 11, t11 vertebra) are similar.
Use of PRA in a matcher

- **Intuition**
  Mapping suggestions with a linguistic similarity vector close to the linguistic similarity vector of a PRA mapping are more likely to be correct suggestions.

- **pmPRA** (Pattern Matcher with PRA)
  - **Strategy**
    - Compute a linguistic similarity vector for each PRA mapping.
    - For each mapping suggestion, we *augment* its similarity value according to the number of PRA mappings within its *neighborhood*. 
Use of PRA in a matcher

- For example
  - Given a suggestion $A$, suppose there are 4 PRA mappings within its neighborhood.

<table>
<thead>
<tr>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Neighborhood Radius = 0.1</td>
</tr>
<tr>
<td>2. Augment per count = 0.06</td>
</tr>
</tbody>
</table>

Original Similarity Value: 0.4 $\Rightarrow$ New Similarity Value: $0.64 (=0.4 + 4 \times 0.06)$
Use of PRA in a matcher

<table>
<thead>
<tr>
<th>Case</th>
<th>RA</th>
<th>PRA</th>
<th>TM</th>
<th>BAMBO</th>
<th>nmpPRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>0.65/1/0</td>
<td>0.65/0.65</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.6</td>
<td>0.4/1/0.65</td>
<td>0.6/1/0.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>4</td>
<td>4</td>
<td>0.4</td>
<td>0.55/0.75/0.65/0.75</td>
<td>0.55/0.75/0.65/0.75</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.75/0.75/0.75/0.75</td>
<td>0.75/0.75/0.75/0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.75/0.65/0.65/0.65</td>
<td>0.75/0.65/0.65/0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rep5</td>
<td>7</td>
<td>4</td>
<td>0.4</td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td></td>
</tr>
<tr>
<td>rep2</td>
<td>27</td>
<td>14</td>
<td>0.4</td>
<td>0.65/0.35/0.31/0.36</td>
<td>0.65/0.35/0.31/0.36</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.55/0.35/0.35/0.35</td>
<td>0.55/0.35/0.35/0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.55/0.55/0.55/0.55</td>
<td>0.55/0.55/0.55/0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rep1</td>
<td>27</td>
<td>12</td>
<td>0.4</td>
<td>0.55/0.35/0.35/0.35</td>
<td>0.55/0.35/0.35/0.35</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.55/0.35/0.35/0.35</td>
<td>0.55/0.35/0.35/0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.55/0.55/0.55/0.55</td>
<td>0.55/0.55/0.55/0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anatomy</td>
<td>1523</td>
<td>985</td>
<td>0.4</td>
<td>0.65/0.55/0.55/0.55</td>
<td>0.65/0.55/0.55/0.55</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.55/0.55/0.55/0.55</td>
<td>0.55/0.55/0.55/0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.55/0.55/0.55/0.55</td>
<td>0.55/0.55/0.55/0.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 4. Using the PRA in a matcher (precision/recall/f1-measure/call-out, etc.)*
Use of PRA in a matcher

- Result Analysis
  - For the small datasets, the correct suggested mappings already had high similarity values, and the missed correct mappings had no shared linguistic pattern with PRA mappings.
  - For the Anatomy dataset, the pmPRA has lower or equal precision. Recall increased for high thresholds and decreased for low thresholds.
    - New correct mappings were found.
    - For low thresholds also new wrong mappings were found.
3. Use of PRA in the filter step
Use of PRA in the filter step

- **fPRA**  (Filter with PRA)
  - **Strategy**
    - Implant PRA mappings in the final result. Any suggestion contradicting with PRA mappings will be filtered out.

- **dtfPRA**  (Double Threshold Filter with PRA)
  - **Strategy**
    - Similar to SAMBOdtf. Use a consistent group in the PRA to filter the suggestions between upper threshold and low threshold.
Use of PRA in the filter step

- **pfPRA** (Pattern Filter with PRA)

- **Strategy**
  1. Cluster all suggestions according to their linguistic similarity vectors using expectation-maximization algorithm.
  2. Assign every PRA mapping to the cluster with the nearest cluster center.
Use of PRA in the filter step

- **Strategy (continued..)**
  
  1. For each cluster, calculate the *average distance* \((\text{AvgDis})\) of PRA mappings to their cluster center.
  
  2. Finally, only suggestions with distance to the cluster center smaller or equal than \(\text{AvgDis}\) will be kept. Otherwise, discarded.
Use of PRA in the filter step (1)

<table>
<thead>
<tr>
<th>Case</th>
<th>RA</th>
<th>PRA</th>
<th>TM</th>
<th>SAMBO</th>
<th>PRA</th>
<th>pPRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>0.6/10.87</td>
<td>0.66/15.37</td>
<td>10.78/0.890.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>0.6/14.87</td>
<td>0.87/10.87</td>
<td>10.78/0.890.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>4</td>
<td>0.4</td>
<td>0.5/0.75/0.75</td>
<td>0.5/0.75/0.75</td>
<td>0.5/0.75/0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>0.5/0.75/0.75</td>
<td>0.5/0.75/0.75</td>
<td>0.5/0.75/0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>4</td>
<td>0.4</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>15.84/0.02/0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>12</td>
<td>77</td>
<td>14</td>
<td>0.4</td>
<td>0.6/0.6/0.91/0.93</td>
<td>0.54/0.86/0.91/0.93</td>
<td>10.32/0.55/0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td></td>
<td>0.5/0.96/0.92/0.93</td>
<td>0.5/0.96/0.92/0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
<td>13</td>
<td>0.4</td>
<td>0.6/0.6/0.9/0.93</td>
<td>0.54/0.87/0.9/0.93</td>
<td>0.45/0.91/0.89/0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td></td>
<td>0.5/0.92/0.95/0.91</td>
<td>0.5/0.92/0.95/0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>14</td>
<td>27</td>
<td>13</td>
<td>0.4</td>
<td>0.6/0.6/0.9/0.93</td>
<td>0.54/0.87/0.9/0.93</td>
<td>0.45/0.91/0.89/0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td></td>
<td>0.5/0.92/0.95/0.91</td>
<td>0.5/0.92/0.95/0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td></td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
</tbody>
</table>

Table 1. Using the PRA during the filter phase - 1 (preokedochemical/mixedochemical1).
Use of PRA in the filter step (1)

Result Analysis

- fPRA always has equal or higher precision and recall than SAMBO.

- pfPRA always has equal or higher precision than fPRA and SAMBO.

- pfPRA always has equal or lower recall than SAMBO.

- Some correct suggestions are filtered out because they have no similar linguistic pattern to PRA mappings.
Use of PRA in the filter step (2)

<table>
<thead>
<tr>
<th>Case</th>
<th>RA</th>
<th>PRA</th>
<th>7%</th>
<th>SAMBO-QM</th>
<th>dPRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>4</td>
<td>2</td>
<td>0.4</td>
<td>0.66/1/0.8/1</td>
<td>1/1/1/1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>0.9/1/0.8/1</td>
<td>1/1/1/1</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>4</td>
<td>0.4</td>
<td>0.45/0.62/0.52/0.62</td>
<td>0.54/0.75/0.63/0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.71/0.62/0.66/0.62</td>
<td>0.75/0.75/0.75/0.50</td>
</tr>
<tr>
<td>max</td>
<td>7</td>
<td>4</td>
<td>0.4</td>
<td>1/1/1/1</td>
<td>1/1/1/1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.6</td>
<td>1/1/1/1</td>
<td>1/1/1/1</td>
</tr>
<tr>
<td>npr</td>
<td>27</td>
<td>14</td>
<td>0.4</td>
<td>0.85/0.92/0.92/0.92</td>
<td>0.86/0.92/0.92/0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.85/0.92/0.92/0.92</td>
<td>0.86/0.92/0.92/0.92</td>
</tr>
<tr>
<td>npa</td>
<td>27</td>
<td>15</td>
<td>0.4</td>
<td>0.85/0.92/0.87/0.92</td>
<td>0.86/0.92/0.86/0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.92/0.86/0.90/0.85</td>
<td>0.92/0.92/0.92/0.85</td>
</tr>
<tr>
<td>Amyot</td>
<td>1513</td>
<td>332</td>
<td>0.4</td>
<td>0.84/0.84/0.84/0.84</td>
<td>0.86/0.87/0.87/0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.89/0.84/0.86/0.84</td>
<td>0.90/0.87/0.88/0.64</td>
</tr>
</tbody>
</table>

*Table 6. Using the PRA during the filter phase - 2 (precision/recall/F-measure/recall/PRA).*
Use of PRA in the filter step (2)

- Result Analysis
  - dtfPRA always has equal or higher recall than SAMBOdtf.
  - For lower threshold 0.6, dtfPRA always has equal or higher precision than SAMBOdtf.
  - For lower threshold 0.4, dtfPRA always has equal or higher precision than SAMBOdtf, except for dataset ear and eye.
  - For dataset ear and eye, the consistent group of dtfPRA is much smaller than the consistent group of SAMBOdtf.
4. Influence of size of PRA
Use of PRA-Full vs PRA-Half

<table>
<thead>
<tr>
<th>Strategy</th>
<th>1%</th>
<th>PRA-F</th>
<th>PRA-H</th>
<th>NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>mgPRA</td>
<td>0.0780/0.570/0.32</td>
<td>0.080/0.550/0.32</td>
<td>331</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>0.080/0.570/0.32</td>
<td>0.080/0.550/0.32</td>
<td>331</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>0.050/0.520/0.35</td>
<td>0.050/0.500/0.36</td>
<td>281</td>
<td>30</td>
</tr>
<tr>
<td>mgICA</td>
<td>0.080/0.550/0.32</td>
<td>0.080/0.570/0.30</td>
<td>336</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>0.050/0.540/0.34</td>
<td>0.050/0.560/0.37</td>
<td>311</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>0.050/0.520/0.36</td>
<td>0.050/0.500/0.38</td>
<td>281</td>
<td>38</td>
</tr>
<tr>
<td>pumPRA</td>
<td>0.080/0.530/0.31</td>
<td>0.080/0.520/0.31</td>
<td>396</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>0.070/0.500/0.31</td>
<td>0.070/0.490/0.31</td>
<td>374</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>0.050/0.500/0.30</td>
<td>0.050/0.490/0.30</td>
<td>300</td>
<td>32</td>
</tr>
<tr>
<td>PRA</td>
<td>0.080/0.550/0.36</td>
<td>0.080/0.520/0.34</td>
<td>357</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>0.060/0.570/0.38</td>
<td>0.060/0.580/0.37</td>
<td>348</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>0.050/0.540/0.38</td>
<td>0.050/0.530/0.38</td>
<td>284</td>
<td>30</td>
</tr>
<tr>
<td>pPRA</td>
<td>0.080/0.510/0.32</td>
<td>0.080/0.500/0.31</td>
<td>179</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>0.050/0.540/0.34</td>
<td>0.050/0.530/0.34</td>
<td>175</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>0.040/0.570/0.33</td>
<td>0.040/0.570/0.33</td>
<td>136</td>
<td>34</td>
</tr>
<tr>
<td>JR-PRA</td>
<td>0.080/0.590/0.37</td>
<td>0.080/0.580/0.37</td>
<td>355</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>0.060/0.570/0.35</td>
<td>0.060/0.560/0.38</td>
<td>348</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 7. Accuracy (1333 correct mappings in the EA) with PRA-F (905 mappings) and PRA-H (424 mappings). - (precision/recall/F-measure), n - N indicates the number of correct mappings not in PRA-F found by using PRA-X. NF is the number of mappings in PRA-F not found by the algorithm using PRA-H.
Use of PRA-Full vs PRA-Half

Result Analysis

For larger PRA

- For all strategies, the recall is higher.

- For the preprocessing strategies and pmPRA
  - When threshold is low, the precision is lower.
  - When threshold is high, the precision is higher.

- For the filtering strategies
  - The precision is always equal or higher.
Lessons learned

- PRA in preprocessing leads to fewer suggestions, in most cases to an improvement in precision and in some cases to an improvement in recall.

- Use the linguistic pattern matcher mainly to find new suggestions.

- Always use filter with PRA. The other filter approaches work well when the structure of the source ontologies is well-defined and complete.

- Not so large difference between PRA-based algorithms and SAMBO/SAMBOdtf
  - SAMBO/SAMBOdtf already do well on test cases
  - Anatomy case: all new correct mappings are non-trivial
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Recommending ontology alignment strategies
- Using PRA in ontology alignment
- Current Issues
Current issues

- Systems and algorithms
  - Complex ontologies
  - Use of instance-based techniques
  - Alignment types (equivalence, is-a, …)
  - Complex mappings (1-n, m-n)
  - Connection ontology types – alignment strategies
Current issues

- Evaluations
  - Need for Golden standards
  - Systems available, but not always the alignment algorithms
  - Evaluation measures

- Recommending ’best’ alignment strategies
Further reading

Starting points for further studies
Further reading
ontology alignment

- [http://www.ontologymatching.org](http://www.ontologymatching.org)
  (plenty of references to articles and systems)

- Ontology alignment evaluation initiative: [http://oaei.ontologymatching.org](http://oaei.ontologymatching.org)
  (home page of the initiative)


  (description of the SAMBO tool and overview of evaluations of different matchers)

  (description of the KitAMO tool for evaluating matchers)
Further reading
ontology alignment

  (double threshold filtering technique)

  (recommendation of alignment strategies)

  (PRA in ontology alignment)
Further reading
ontology alignment


(largest overview of systems)