Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges
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)**

<table>
<thead>
<tr>
<th>immune response</th>
</tr>
</thead>
<tbody>
<tr>
<td>i- acute-phase response</td>
</tr>
<tr>
<td>i- anaphylaxis</td>
</tr>
<tr>
<td>i- antigen presentation</td>
</tr>
<tr>
<td>i- antigen processing</td>
</tr>
<tr>
<td>i- cellular defense response</td>
</tr>
<tr>
<td>i- cytokine metabolism</td>
</tr>
<tr>
<td>i- cytokine biosynthesis</td>
</tr>
<tr>
<td>synonym cytokine production</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>p- regulation of cytokine biosynthesis</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>

...|

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
      - 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
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
      - regulation of cytokine biosynthesis
      - ... (omitted)
  - ... (omitted)
- B-cell activation
  - B-cell differentiation
  - B-cell proliferation
  - cellular defense response
  - ... (omitted)
- T-cell activation
  - activation of natural killer cell activity
  - ... (omitted)

**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
Defining the relations between the terms in different ontologies
Ontology Alignment

- Ontology alignment
- **Ontology alignment strategies**
- Evaluation of ontology alignment strategies
- Ontology alignment challenges
An Alignment Framework
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` $\rightarrow$ `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
Example matchers

- Instance-based
- Use life science literature as instances
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.
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

$$sim(C_1, C_2) = \frac{n_{NBC2}(C_1, C_2) + n_{NBC1}(C_2, C_1)}{n_D(C_1) + n_D(C_2)}$$
Matcher Strategies

- Strategies based linguistic matching
- 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>String based strategies</th>
<th>Structure-based strategies</th>
<th>Constraint-based strategies</th>
<th>Instance-based strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgreementMaker</td>
<td>Similarities, Edit Distance, TF-IDF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ALIN</td>
<td>Similarities APP, WS4J APP</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AML</td>
<td>Jaccard, I Sub</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Anchor-Flood</td>
<td>Jaro-Winkler</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>AOAS</td>
<td>Jaro-Winkler</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>AOT, AOTL</td>
<td>Edit Distance, Block Distance, SLM-Winkler, Jaro-Winkler, Smith-Winkler, Needleman-Wunsch</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>AROMA</td>
<td>Jaro-Winkler</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ASMOV</td>
<td>Edit Distance</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>BLOOMS</td>
<td>Jaccard, Exact Match, Lin, Jaro-Winkler</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CIDER-CL</td>
<td>Soft TF-IDF, Jaro-Winkler</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CODI</td>
<td>Edit Distance, Jaro-Winkler, Cosine, Smith-Waterman, Jaccard, Overtag coefficient</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>COMMAND</td>
<td>UMBC similarity Model</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CroMatcher</td>
<td>N-Gram, TF-IDF</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CSA</td>
<td>Edit Distance, Wu-Palmer, TF-IDF</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DKP-AOM, DKP-AOM-Lite</td>
<td>Similarities APP</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DSSim</td>
<td>Jaccard, Jaro-Winkler</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eff2Match</td>
<td>Exact Match, TF-IDF</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Falcon-AO</td>
<td>I Sub, TF-IDF</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FCA-Map</td>
<td>Exact Match</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GeRoMeSuite+SMB</td>
<td>Edit Distance, Jaro-Winkler, I Sub, Soft TF-IDF, SecondString Library</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>GMap</td>
<td>Edit Distance, TF-IDF</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GOMMA, GOMMA-bk</td>
<td>Exact Match, N-gram</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hertuda</td>
<td>Damerau-Levenshtein²</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HotMatch</td>
<td>Damerau-Levenshtein²</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IAMA</td>
<td>Edit Distance</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 8 Matching strategies in the participating systems - 2

<table>
<thead>
<tr>
<th>System</th>
<th>String-based strategies</th>
<th>Structure-based strategies</th>
<th>Constraint-based strategies</th>
<th>Instance-based strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>JarvisOM</td>
<td>Cosine, WuPalmer, Lin, N-gram</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KOSIMap</td>
<td>SimMetrics APP, Degree of commonality coefficient</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Lily</td>
<td>Edit Distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LogMap</td>
<td>1-Sub</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>LPHOM</td>
<td>1-Sub, Mongue-Elian, 3-Gram</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYAM++</td>
<td>SOFT TF-IDF, Jaccard</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaasMatch</td>
<td>Cosine, Edit-Distance, Jaccard, 3-Gram, Longest Common Substring</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>MapSSS</td>
<td>Edit-Distance, Choice based on [10]</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>NBJLM</td>
<td>Set of words-level</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODGOMS</td>
<td>Longest Common Subsequence, SMOA, TF-IDF</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optima+</td>
<td>Lin, Smith Waterman, Needleman-Wunsch, Inverse Edit-Distance</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior+</td>
<td>Edit-Distance</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIMOM</td>
<td>Edit-Distance, Cosine</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>RSDLWB</td>
<td>Jaccard, Substring</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAMBO, SAMBodtf</td>
<td>Edit-Distance, 3-Gram</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>ServOMap</td>
<td>Edit-Distance, 1-Sub, Q-Gram, TF-IDF, Mongue-Elian, Jaccard</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOBOM</td>
<td>1-Sub</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StringsAuto</td>
<td>Choice based on [10]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TaxoMap</td>
<td>Lin, 3-gram</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| TOAST        | ✓
| WeSeE        | Edit-Distance, TF-IDF                                 |                            |                             |                           |
| WikiMatch    | Jaccard                                               |                            |                             |                           |
| X-SOM        | Edit Distance, Jaro                                   | ✓                          |                            | ✓                         |
| XMap         | Edit distance, Jaro-Winkler, N-gram, Jaccard, Cosine  | ✓                          | ✓                           |                           |
| YAM++        | Tversky⁷, TF-IDF                                      | ✓                          |                            | ✓                         |

<table>
<thead>
<tr>
<th>System</th>
<th>Background knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgreementMaker</td>
<td>✓ ✓ - - - - ✓</td>
</tr>
<tr>
<td>AILIN</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>AML</td>
<td>✓ ✓ - - - - -</td>
</tr>
<tr>
<td>Anchor Flood</td>
<td>- ✓ - - - - -</td>
</tr>
<tr>
<td>AOAS</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>AOT, AOTL</td>
<td>- ✓ - - - - -</td>
</tr>
<tr>
<td>ASMOV</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>COMMAND</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>CroMatcher</td>
<td>- ✓ - - - - -</td>
</tr>
<tr>
<td>CSA</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>DIP-AOM</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>DSSim</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>EIDIMatch</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>GOMMA</td>
<td>✓ ✓ - - - - -</td>
</tr>
<tr>
<td>GetRelMeSuite+SMR</td>
<td>- - ✓ - - - -</td>
</tr>
<tr>
<td>Hormarch</td>
<td>✓ ✓ - - - - -</td>
</tr>
<tr>
<td>JarvisOM</td>
<td>- - - - - ✓ -</td>
</tr>
<tr>
<td>IAMA</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>Lity</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>LogMapBio</td>
<td>- - ✓ - - - -</td>
</tr>
<tr>
<td>LYAM+1</td>
<td>- ✓ - - - - -</td>
</tr>
<tr>
<td>MaasMarch</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>MapSSS</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>NBLSM</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>Optima1</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>RMOM</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>RSDIMWB</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>SAMBO</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>ServOMap</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>TaxoMap</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>TOAST</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>WeSeE</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>WikiMatch</td>
<td>- - ✓ - - - -</td>
</tr>
<tr>
<td>XMap</td>
<td>✓ - - - - - -</td>
</tr>
<tr>
<td>X GOM</td>
<td>- - - - - - -</td>
</tr>
<tr>
<td>YAM++</td>
<td>- - - - - - -</td>
</tr>
</tbody>
</table>

Note: The background knowledge sources are as follows:
- UMLS
- Uberon
- BioPortal
- MeSH
- FMA
- WordNet
- Other
- API lines, Wikipedia
- Big Huge Thesaurus
- Apache Lucene
- Apache Lucene
- Web search (Google)
- BabelNet
- Google
- Wiki Pages
- DBpedia
- Apache Lucene
- Apache Lucene
- Microsoft Bing Search
- JfreeWetSearch
- WikiPedia
- Google
- Apache Lucene
Combinations
Combination Strategies

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

- Threshold filtering

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

\[
\begin{align*}
(2, B) \\
(3, F) \\
(6, D) \\
(4, C) \\
(5, C) \\
(5, E) \\
\vdots
\end{align*}
\]

\[
\text{suggest} \quad \text{discard}
\]
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)

\[
\begin{align*}
\text{upper-th} & : (2, B) \\
& : (3, F) \\
& : (6, D) \\
& : (4, C) \\
& : (5, C) \\
\text{lower-th} & : (5, E) \\
& \ldots \ldots
\end{align*}
\]
Example alignment system
SAMBO – matchers, combination, filter
Example alignment system
SAMBO – suggestion mode
<table>
<thead>
<tr>
<th>Systems</th>
<th>Preprocessing</th>
<th>Matching</th>
<th>Combination</th>
<th>Filtering</th>
<th>Debugging</th>
<th>User interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgreementMaker</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>AILIN</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AML, AML_bk</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Anchor-Flood</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AOAS</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AOT, AOTL</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>AROMA</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ASMOV</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>BLOOMS</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>CIDER-CL</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CODI</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>COMMAND</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CroMatcher</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CSA</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>DKP-AOM, DKP-AOM-Lite</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>DSSim</td>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>EMDMatch</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Falcon-AO</td>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>FCA-Map</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GeBoMeSuite-CA</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GMap</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>GOMMA, GOMMA-bk</td>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>Herraduda</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HotMatch</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>IAMA</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>D</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓*</th>
</tr>
</thead>
<tbody>
<tr>
<td>JarvisOM</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>KOSIMap</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>Lity</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓*</td>
</tr>
<tr>
<td>LogMap, LogMapBio, LogMapC, LogMapLite</td>
<td>D,R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓*</td>
</tr>
<tr>
<td>LPHOM</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LYAM++</td>
<td>D</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MaasMatch</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MapSSS</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NBJLM</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>OGOMS</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Optimat+</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prior+</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>RIMOM</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>RSDLWB</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓*</td>
</tr>
<tr>
<td>SAMBO, SAMBOostrf</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓*</td>
</tr>
<tr>
<td>ServOMap(L), ServOMBI</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SOBOM</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>StringsAuto</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TaxoMap</td>
<td>D,R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TOAST</td>
<td>-</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>WeSeE</td>
<td>D</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>WildMatch</td>
<td>D</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>X-SOM</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>XMap, XMAPGen, XMAPSig</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>YAM++</td>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
</tbody>
</table>
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges
Evaluation measures

- Precision:
  \[
  \frac{\text{# correct mapping suggestions}}{\text{# mapping suggestions}}
  \]

- Recall:
  \[
  \frac{\text{# correct mapping suggestions}}{\text{# correct mappings}}
  \]

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

http://oaei.ontologymatching.org/
OAEI

- Since 2004, Evaluation of systems
- Different tracks (2020)
  - Ontologies
    - Anatomy, conference, large biomedical ontologies, disease and phenotype, biodiversity and ecology
    - Multilingual: multifarm (9 languages)
    - Complex
    - Interactive
  - Instance matching and link discovery
  - Knowledge graphs
OAEI

Evaluation measures

- Precision/recall/f-measure
- recall of non-trivial mappings
- full / partial golden standard
OAEI 2019

- 12 systems

Anatomy:

- best system $f=0.943$, $p=0.95$, $r=0.936$, $r+=0.832$, 76 seconds (42s in 2018)
- 4 systems produce coherent mappings (5 in 2018)
OAEI Anatomy Track 2007-2016*

- Components
  - Almost all systems implement preprocessing, matchers, combination, filtering components
  - Debugging component and GUI rarely implemented

- Matching strategies
  - Variety of string-based strategies
  - Most often string and structured-based strategies

- Use of background knowledge
  - Almost all systems use sources of background knowledge

Complementary evaluation

Alignment cubes

- Interactive visualization of alignments
- Region-level, mapping level
- Missing mappings
- Often found mappings

http://www.ida.liu.se/~patla00/research/AlignmentCubes/
Alignment cubes
Ontology Alignment

- Ontology alignment
- Ontology alignment strategies
- Evaluation of ontology alignment strategies
- Ontology alignment challenges
Challenges

- Large-scale matching evaluation
- Efficiency of matching techniques
  - parallellization
  - distribution of computation
  - approximation of matching results (not complete)
  - modularization of ontologies
  - optimization of matching methods
Challenges

- Matching with background knowledge
  - partial alignments
  - reuse of previous matches
  - use of domain-specific corpora
  - use of domain-specific ontologies

- Matcher selection, combination and tuning
  - recommendation of algorithms and settings
Challenges

- User involvement
  - visualization
  - user feedback

- Explanation of matching results
- Social and collaborative matching
- Alignment management: infrastructure and support
Further reading

Starting points for further studies
Further reading
ontology alignment

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

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


Further reading

ontology alignment

Systems at LiU / IDA / ADIT


Further reading
ontology alignment

  (double threshold filtering technique)

  (recommendation of alignment strategies)

  (use of partial alignments in ontology alignment)
Further reading
ontology alignment

User Involvement


Ontology Completion and Debugging
Defects in ontologies

- Syntactic defects
  - E.g. wrong tags or incorrect format

- Semantic defects
  - E.g. unsatisfiable concepts, incoherent and inconsistent ontologies

- Modeling defects
  - E.g. wrong or missing relations
Example - incoherent ontology

Example: DICE ontology

- \textbf{Brain} \subseteq \textbf{CentralNervousSystem} \sqcap \textbf{BodyPart} \sqcap
  \exists \text{systempart.NervousSystem} \sqcap \exists \text{region.HeadAndNeck} \sqcap
  \forall \text{region.HeadAndNeck}

A brain is a central nervous system and a body part which has a system part that is a nervous system and that is in the head and neck region.

- \textbf{CentralNervousSystem} \subseteq \textbf{NervousSystem}

A central nervous system is a nervous system.

- \textbf{BodyPart} \subseteq \neg \text{NervousSystem}

Nothing can be at the same time a body part and a nervous system.
Example - inconsistent ontology

- **Example from Foaf:**
  - Person(timbl)
  - Homepage(timbl, http://w3.org/)
  - Homepage(w3c, http://w3.org/)
  - Organization(w3c)
  - InverseFunctionalProperty(Homepage)
  - DisjointWith(Organization, Person)

- **Example from OpenCyc:**
  - ArtifactualFeatureType(PopulatedPlace)
  - ExistingStuffType(PopulatedPlace)
  - DisjointWith(ExistingObjectType, ExistingStuffType)
  - ArtifactualFeatureType \subseteq ExistingObjectType
Example - missing is-a relations

- In 2008 Ontology Alignment Evaluation Initiative (OAEI) Anatomy track, task 4
  - Ontology MA : Adult Mouse Anatomy Dictionary (2744 concepts)
  - Ontology NCI-A : NCI Thesaurus - anatomy (3304 concepts)
  - 988 mappings between MA and NCI-A
    - 121 missing is-a relations in MA
    - 83 missing is-a relations in NCI-A
Influence of missing structure

- Ontology-based querying.

Medical Subject Headings (MeSH)

All MeSH Categories

- Diseases Category
  - Eye Diseases
    - Scleral Diseases
    - Scleritis

... return 1617 articles
Influence of missing structure

- Incomplete results from ontology-based queries

![PubMed search example](image)

**Medical Subject Headings (MeSH)**

All MeSH Categories
- Diseases Category
  - Eye Diseases
    - Scleral Diseases
    - Scleritis

- Search: "Scleral Diseases" [MeSH]
  - Return 1617 articles
  - Return 695 articles
  - 57% results are missed!
Defects in ontologies and ontology networks

- Ontologies and ontology networks with defects, although often useful, also lead to problems when used in semantically-enabled applications.

→ Wrong conclusions may be derived or valid conclusions may be missed.
Completion and debugging process

- Detection (find candidate defects)
- Validation (real defects)
- Repair (remove wrong, add correct)
Detection

Many approaches

- inspection
- ontology learning or evolution
- using linguistic and logical patterns
  - animals *such as* dogs and cats
- by using knowledge intrinsic to an ontology network
- by using machine learning and statistical methods
Repairing

Definition 1 (Repair) Let $T$ be a TBox and $C$ be the set of all atomic concepts in $T$. Let $M$ and $W$ be finite sets of TBox axioms. Let $Or$ be an oracle that given a TBox axiom returns true or false. A repair for Complete-Debug-Problem $CDP(T, C, Or, M, W)$ is any pair of finite sets of TBox axioms $(A, D)$ such that

(i) $\forall \psi_a \in A: Or(\psi_a) = true$;
(ii) $\forall \psi_d \in D: Or(\psi_d) = false$;
(iii) $(T \cup A) \setminus D$ is consistent;
(iv) $\forall \psi_m \in M: (T \cup A) \setminus D \models \psi_m$;
(v) $\forall \psi_w \in W: (T \cup A) \setminus D \not\models \psi_w$.

Current work usually focuses on debugging or completion, but not both.
Most work on debugging.
Ontology Debugging
Example: an Incoherent Ontology

Consider the following TBox $T^*$, where $A$, $B$ and $C$ are primitive and $A_1, \ldots, A_7$ defined concept names:

\[
\begin{align*}
ax_1 & : A_1 \sqsubseteq \neg A \sqcap A_2 \sqcap A_3 \\
ax_2 & : A_2 \sqsubseteq A \sqcap A_4 \\
ax_3 & : A_3 \sqsubseteq A_4 \sqcap A_5 \\
ax_4 & : A_4 \sqsubseteq \forall s.B \sqcap C \\
ax_5 & : A_5 \sqsubseteq \exists s.\neg B \\
ax_6 & : A_6 \sqsubseteq A_1 \sqcup \exists r.(A_3 \sqcap \neg C \sqcap A_4) \\
ax_7 & : A_7 \sqsubseteq A_4 \sqcap \exists s.\neg B \\
\end{align*}
\]

\[\downarrow\]

DL Reasoner

The ontology is incoherent!
The set of unsatisfiable concepts are: \{ $A_1$, $A_3$, $A_6$, $A_7$ \}.

What are the root causes of these defects?
Explain the Semantic Defects

We need to identify the sets of axioms which are necessary for causing the logic contradictions.

\[
\begin{align*}
ax_1 & : A_1 \models \neg A \sqcap A_2 \sqcap A_3 & ax_2 & : A_2 \models A \sqcap A_4 \\
ax_3 & : A_3 \models A_4 \sqcap A_5 & ax_4 & : A_4 \models \forall s. B \sqcap C \\
ax_5 & : A_5 \models \exists s. \neg B & ax_6 & : A_6 \models A_1 \sqcup \exists r.(A_3 \sqcap \neg C \sqcap A_4) \\
ax_7 & : A_7 \models A_4 \sqcap \exists s. \neg B
\end{align*}
\]

For example, for the unsatisfiable concept “\(A_1\)”, there are two sets of axioms.

\[
\begin{align}
ax_1 & : A_1 \models \neg A \sqcap A_2 \sqcap A_3 \\
ax_2 & : A_2 \models A_4 \sqcap A_4
\end{align}
\]
Minimal Unsatisfiability Preserving Sub-TBoxes (MUPS)

**Definition 1** Let $A$ be a concept which is unsatisfiable in a TBox $\mathcal{T}$. A set $\mathcal{T}' \subseteq \mathcal{T}$ is a minimal unsatisfiability-preserving sub-TBox (MUPS) of $\mathcal{T}$ if

- $A$ is unsatisfiable in $\mathcal{T}'$, and
- $A$ is satisfiable in every sub-TBox $\mathcal{T}'' \subset \mathcal{T}'$.

We will abbreviate the set of MUPS of $\mathcal{T}$ and $A$ by $mups(\mathcal{T}, A)$.

$$mups(\mathcal{T}^*, A_1) = \{\{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\}\}$$

- The MUPS of an unsatisfiable concept imply the solutions for repairing.
  - Remove at least one axiom from each axiom set in the MUPS
Example

Possible ways of repairing all the unsatisfiable concepts in the ontology:

\[ mups(T^*, A_1) = \{\{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\}\} \]
\[ mups(T^*, A_3) = \{\{ax_3, ax_4, ax_5\}\} \]
\[ mups(T^*, A_6) = \{\{ax_1, ax_2, ax_4, ax_6\}, \{ax_1, ax_3, ax_4, ax_5, ax_6\}\} \]
\[ mups(T^*, A_7) = \{\{ax_4, ax_7\}\} \]

How to represent all these possibilities?
Definition 2 Let $\mathcal{T}$ be an incoherent TBox. A TBox $\mathcal{T}' \subseteq \mathcal{T}$ is a minimal incoherence-preserving sub-TBox (MIPS) of $\mathcal{T}$ if

- $\mathcal{T}'$ is incoherent, and
- every sub-TBox $\mathcal{T}'' \subseteq \mathcal{T}'$ is coherent.

\[
mups(\mathcal{T}^*, A_1) = \{\{ax_1, ax_2\}, \{ax_1, ax_3, ax_4, ax_5\}\}
mups(\mathcal{T}^*, A_3) = \{\{ax_3, ax_4, ax_5\}\}
mups(\mathcal{T}^*, A_6) = \{\{ax_1, ax_2, ax_4, ax_6\},\}
\{ax_1, ax_3, ax_4, ax_5, ax_6\}\}
mups(\mathcal{T}^*, A_7) = \{\{ax_3, ax_7\}\}
\]

We will abbreviate the set of MIPS of $\mathcal{T}$ by $\text{mips}(\mathcal{T})$. For $\mathcal{T}^*$ we get three MIPS:

\[
\text{mips}(\mathcal{T}^*) = \{\{ax_1, ax_2\}, \{ax_3, ax_4, ax_5\}, \{ax_4, ax_7\}\}
\]

A possible repairing is $\{ax_i\} \cup \{ax_j\} \cup \{ax_k\}$, where

- $ax_i \in \{ax_1, ax_2\}$
- $ax_j \in \{ax_3, ax_4, ax_5\}$
- $ax_k \in \{ax_4, ax_7\}$
Completing the is-a structure of ontologies
Example

Repairing actions:

\{\text{Endocarditis} \sqsubseteq \text{PathologicalPhenomenon}, \text{GranulomaProcess} \sqsubseteq \text{NonNormalProcess}\}
\{\text{Carditis} \sqsubseteq \text{CardioVascularDisease}, \text{GranulomaProcess} \sqsubseteq \text{PathologicalProcess}\}
\{\text{Carditis} \sqsubseteq \text{Fracture}, \text{GranulomaProcess} \sqsubseteq \text{NonNormalProcess}\}
Description logic EL

- Concepts

<table>
<thead>
<tr>
<th>Atomic concept</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal concept</td>
<td>$\top$</td>
</tr>
<tr>
<td>Intersection of concepts</td>
<td>$C \cap D$</td>
</tr>
<tr>
<td>Existential restriction</td>
<td>$\exists r.C$</td>
</tr>
</tbody>
</table>

- Terminological axioms: equivalence and subsumption
Generalized Tbox Abduction Problem – GTAP(T, C, Or, M)

- Given
  - T - a Tbox in EL
  - C - a set of atomic concepts in T
  - M = \{A_i \subseteq B_i\}_{i=1..n} \text{ and } \forall i:1..n: A_i, B_i \in C
  - Or: \{C_i \subseteq D_i | C_i, D_i \in C\} \rightarrow \{\text{true, false}\}

- Find
  - S = \{E_i \subseteq F_i\}_{i=1..k} \text{ such that }
    \forall i:1..k: E_i, F_i \in C \text{ and } \text{Or}(E_i \subseteq F_i) = \text{true}
    \text{ and } T \cup U S \text{ is consistent and } T \cup U S \models M
GTAP - example

\[
C = \{ \text{GranulomaProcess, CardioVascularDisease, PathologicalPhenomenon, Fracture, Endocarditis, Carditis, InflammationProcess, PathologicalProcess, NonNormalProcess} \}
\]

\[
T = \{ \text{GranulomaProcess} \sqsubseteq T, \text{hasAssociatedProcess} \sqsubseteq T \times T, \\
\text{CardioVascularDisease} \sqsubseteq \text{PathologicalPhenomenon, Fracture} \sqsubseteq \text{PathologicalPhenomenon,} \\
\exists \text{hasAssociatedProcess.PathologicalProcess} \sqsubseteq \text{PathologicalPhenomenon,} \\
\text{Endocarditis} \sqsubseteq \text{Carditis}, \text{Endocarditis} \sqsubseteq \exists \text{hasAssociatedProcess.InflammationProcess,} \\
\text{PathologicalProcess} \sqsubseteq \text{NonNormalProcess} \}
\]

\[
M = \{ \text{Endocarditis} \sqsubseteq \text{PathologicalPhenomenon, GranulomaProcess} \sqsubseteq \text{NonNormalProcess} \}
\]
Preference criteria

- There can be many solutions for GTAP
Preference criteria

- There can be many solutions for GTAP

Not all are equally interesting.
More informative

- Let $S$ and $S'$ be two solutions to $\text{GTAP}(\mathbf{T}, \mathbf{C}, \text{Or}, M)$. Then,
  - $S$ is more informative than $S'$ iff $\mathbf{T} \cup S \models S'$ but not $\mathbf{T} \cup S' \models S$
  - $S$ is equally informative as $S'$ iff $\mathbf{T} \cup S \models S'$ and $\mathbf{T} \cup S' \models S$
More informative

- 'Blue' solution is more informative than 'green' solution
Semantic maximality

A solution $S$ to $\text{GTAP}(T,C,\text{Or},M)$ is semantically maximal iff there is no solution $S'$ which is more informative than $S$. 
Subset minimality

- A solution $S$ to $\text{GTAP}(T,C,Or,M)$ is subset minimal iff there is no proper subset $S'$ of $S$ that is a solution.
Combining with priority for semantic maximality

- A solution $S$ to $\text{GTAP}(T, C, \text{Or}, M)$ is maxmin optimal iff $S$ is semantically maximal and there is no other semantically maximal solution that is a proper subset of $S$. 
Combining with priority for subset minimality

- A solution $S$ to GTAP($T, C, Or, M$) is minmax optimal iff $S$ is subset minimal and there is no other subset minimal solution that is more informative than $S$. 
Combining with equal preferences

- A solution $S$ to $\text{GTAP}(T,C,\text{Or},M)$ is skyline optimal iff there is no other solution that is a proper subset of $S$ and that is equally informative than $S$.
  - All subset minimal, minmax optimal and maxmin optimal solutions are also skyline optimal solutions.
  - Semantically maximal solutions may or may not be skyline optimal.
Preference criteria - conclusions

- In practice it is not clear how to generate maxmin or semantically maximal solutions (the preferred solutions)
- Skyline optimal solutions are the next best thing and are easy to generate
Approach

- **Input**
  - Normalized EL - TBox
  - Set of missing is-a relations (correct according to the domain)

- **Output** – a skyline-optimal solution to GTAP

- **Iteration of three main steps:**
  - Creating solutions for individual missing is-a relations
  - Combining individual solutions
  - Trying to improve the result by finding a solution which introduces additional new knowledge (more informative)
Intuition 1

Source set

Prokaryote

ArchaebacteriaCell

Target set

Organism

MulticellularOrganism

Alga

AdultOrganism

JuvenileOrganism

EmbryonicOrganism
Intuitions 2/3
Example – repairing single is–a relation

Endocarditis ⬈ PathologicalPhenomenon
Endocarditis ⬈ Fracture
Endocarditis ⬈ CardioVascularDisease
Carditis ⬈ PathologicalPhenomenon
Carditis ⬈ Fracture
Carditis ⬈ CardioVascularDisease
InflammationProcess ⬈ PathologicalProcess
Example – repairing single is–a relation

GranulomaProcess ∈ NonNormalProcess
GranulomaProcess ∈ PathologicalProcess
Algorithm - Repairing multiple is-a relations

- Combine solutions for individual missing is-a relations
- Remove redundant relations while keeping the same level of informativeness
- Resulting solution is a skyline optimal solution

{InflammationProcess \sqsubseteq PathologicalProcess, Carditis \sqsubseteq CardioVascularDisease, GranulomaProcess \sqsubseteq PathologicalProcess}
Algorithm – improving solution

- Solution S from previous step may contain relations which are not derivable from the ontology.
- These can be seen as new missing is-a relations.
- We can solve a new GTAP problem: GTAP(T U S, C, Or, S)
Example – improving solutions

GranulomaProcess ⊆ InflammationProcess

{InflammationProcess ⊆ PathologicalProcess,
Carditis ⊆ CardioVascularDisease,
GranulomaProcess ⊆ InflammationProcess}
Algorithm properties

- Sound
- Skyline optimal solutions
Experiments

Two use-cases

- Case 1: given missing is-a relations
  AMA and a fragment of NCI-A ontology – OAEI 2013
  - AMA (2744 concepts) – 94 missing is-a relations
    → 3 iterations, 101 in repairing (47 additional new knowledge)
  - NCI-A (3304 concepts) – 58 missing is-a relations
    → 3 iterations, 54 in repairing (10 additional new knowledge)

- Case 2: no given missing is-a relations
  Modified BioTop ontology
  - Biotop (280 concepts, 42 object properties)
    randomly choose is-a relations and remove them: 47 ‘missing’
    → 4 iterations, 41 in repairing (40 additional new knowledge)
Further reading

Starting points for further studies
Further reading ontology debugging

Debugging and Completing Ontologies


Debugging Ontologies

Further reading
ontology debugging

Completing ontologies

