Structure-based Filtering for Ontology Alignment

Bi Chen, He Tan, Patrick Lambrix

Department of Computer and Information Science
Linköpings universitet

Outline

- Introduction
  - Ontology alignment
  - Ontology alignment systems
  - Structure-based alignment algorithm
  - Filtering algorithm
- Structure-based filtering
  - Basic ideas
  - Algorithms
  - Implementation
- Evaluation
- Conclusion

Ontology Alignment

- Many ontologies have been developed

→ Many of them have overlapping information

- Use of multiple ontologies
  e.g. custom-specific ontology + standard ontology
- Bottom-up creation of ontologies
  experts can focus on their domain of expertise

→ Important to know the inter-ontology relationships

Ontology Alignment

GENE ONTOLOGY (GO)

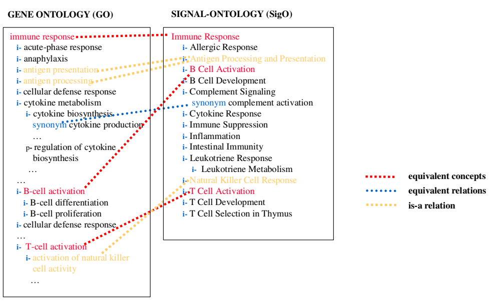
immune response
i- acute-phase response
i- anaphylaxis
i- antigen presentation
i- antigen processing
i- cellular defense response
i- cytokine metabolism
i- cytokine biosynthesis
synonym cytokine production
…
p- regulation of cytokine biosynthesis
…
…
i- B-cell activation
i- B-cell differentiation
i-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

SIGNAL-ONTOLOGY (SigO)

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

- To define the relations between the terms in different ontologies

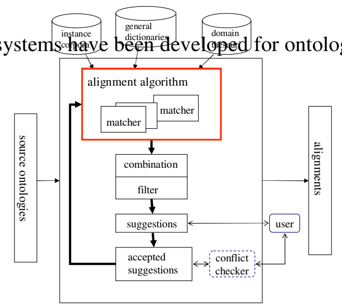
Ontology Alignment



- To define the relations between the terms in different ontologies

Ontology Alignment Systems

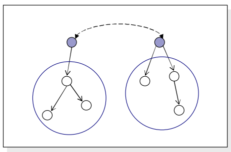
- Many systems have been developed for ontology alignment



Lambrix P, Tan H, 'SAMBO – A System for Aligning and Merging Biomedical Ontologies', Journal of Web Semantics, Special issue on Semantic Web for the Life Sciences. 2006.

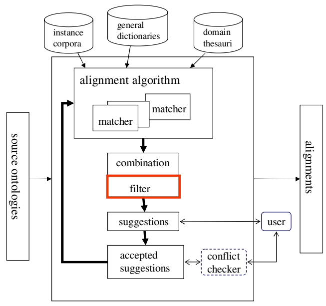
Structure-based Alignment Algorithm

- The intuition is that the similarity of concepts is influenced by their structural environment.
  - often requires previous similarity results
  - intuitively appealing, but poor results (in our experiments)



7

Filtering Algorithm



8

Filtering Algorithm

- Threshold algorithm

  e.g. retains pairs with the similarity above or equal to a threshold

O_1 : {1, 2, 3, 4, 5}
O_2 : {A, B, C, D, E, F}

sim ↓
(2, B)
(3, F) ↗ suggest
(6, D)
(4, C)
th - - - - - - - - - -
(5, C)
(5, E) ↘ discard
……

9

Filtering Algorithm

- When the threshold is high, the precision is high, but the recall can be low.
- When the threshold decreases, the recall increases, but the precision can decrease.

O_1 : {1, 2, 3, 4, 5}
O_2 : {A, B, C, D, E, F}

(2, B)
(3, F)
(6, D)
(4, C)
th - - - - - - - - - -
(5, C)
(5, E)
th - - - - - - - - - -
……

10

Outline

- Introduction
- Structure-based filtering
  - Basic ideas
  - Algorithms
  - Implementation
- Evaluation
- Conclusion

11

Basic ideas

- Retain the pairs : *sim* ≥ *upper-th*
- Filter out the pairs : *upper-th* > *sim* ≥ *lower-th*
- Build a filter using the structural information

O_1 : {1, 2, 3, 4, 5}
O_2 : {A, B, C, D, E, F}

(2, B)
(3, F)
(6, D)
(4, C)
upper-th - - - - - - - - - -
(5, C)
(5, E)
lower-th - - - - - - - - - -
……

consistent with respect to structure

filter ▷ (5, E)

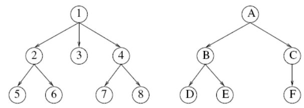
12

Algorithms

- Find a consistent suggestion group from the pairs with similarity values higher than or equal to the upper threshold
- Use the consistent suggestion group to partition the original ontologies
- Filter the pairs with similarity values between the lower and upper thresholds using the partition

Consistent suggestion group

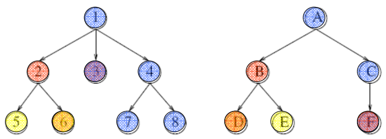
- each concept occurs at most once in the group
- for (A, A') and (B, B'): $A \subset B \Leftrightarrow A' \subset B'$



{ (2, *B*), (*3*, *F*), (*6*, *D*), (*4*, *C*)} is not a CSG
{(*2*, *B*), (*3*, *F*), (*6*, *D*)} is a CSG

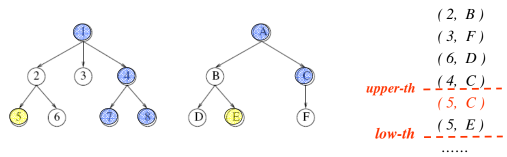
Partition ontologies

CSG: {(*2*, *B*), (*3*, *F*), (*6*, *D*)}



Partition: ({ *5* }, { *E* }), ({ *1*, *4*, *7*, *8* }, { *A*, *C* })

Filtering suggestions



ontology partition: ({ *5* }, { *E* }), ({ *1*, *4*, *7*, *8* }, { *A*, *C* })
→ (*5*, *E*) is retained as a suggestion, but (*5*, *C*) is discarded.

Implementation

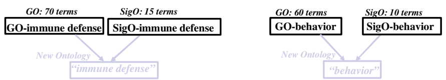
- Find a (maximal) consistent suggestion group
  - A genetic algorithm is used
- Partition original ontologies
  - Based on a binary tree built from ontologies

Outline

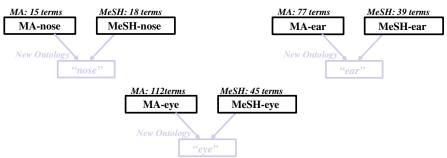
- Introduction
- Structure-based filtering
- Evaluation
- Conclusion

Evaluation

- GO vs. SigO

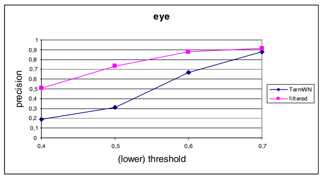


- MA vs. MeSH



Evaluation

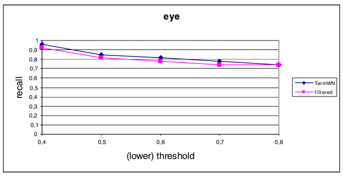
- The precision is increased after filtering.



- a linguistic alignment algorithm using WordNet
- the upper threshold is 0.8

Evaluation

- The recall is constant in most cases after filtering



- a linguistic alignment algorithm using WordNet
- the upper threshold is 0.8

Evaluation

- Time for filtering is only a fraction of the time for the similarity computation

Case	B	ID	nose	ear	eye
TermWN	10.2	35.8	7.7	35.3	60.7
Filter	1.4	1.8	1.0	4.9	5.0

Outline

- Introduction
- Structure-based filtering
- Evaluation
- Conclusion

Conclusion

- Proposed a novel structure-based method for filtering alignment suggestions
- The method gives good results in the experiment.
- In the future, we will investigate the use of the partition to increase the recall.