USING A REFERENCE ONTOLOGY WITH SEMANTIC SIMILARITY FOR ONTOLOGY ALIGNMENT

Valerie Cross
crossv@muohio.edu
Pramit Silwal, David Morell

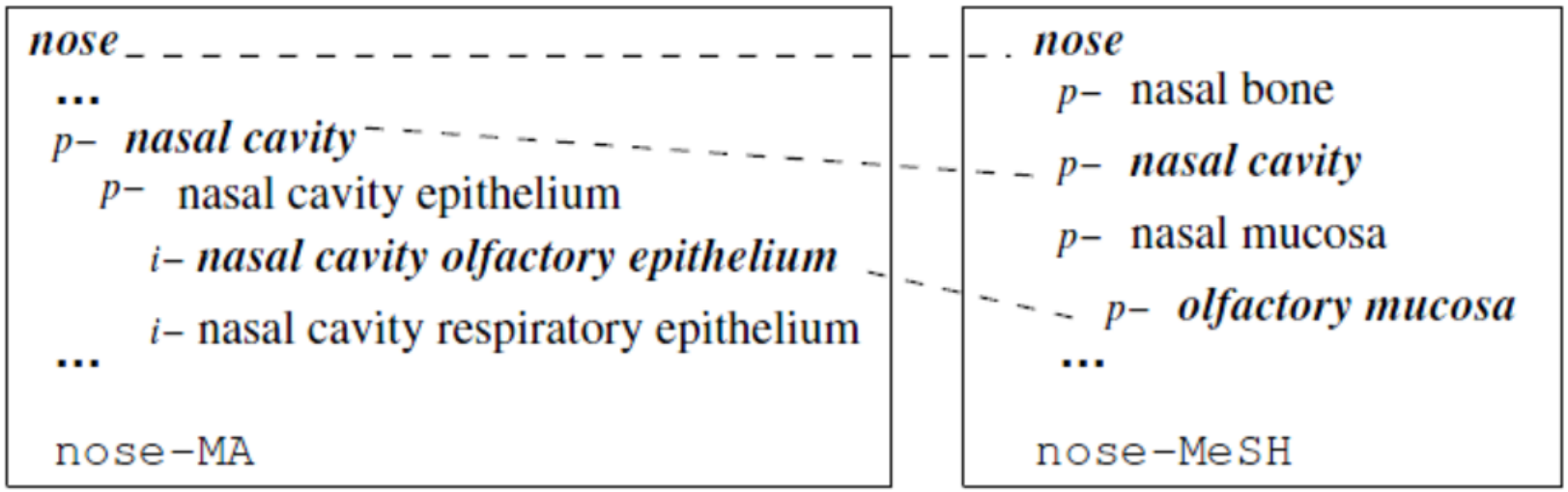
Miami University
Oxford, OH

FOIS 7th International Conference on Formal Ontology in Information Systems July 24 – 27
ICBO 2012
3rd International Conference on Biomedical Ontology July 21 – 25

OVERVIEW

- Ontology Alignment (OA)
- Semantic Similarity (SS) measures
- Recent Approaches Using Reference Ontology
- Reference Ontology + Semantic Similarity
- Initial Experiment and Results
- Future Work

ONTOLOGY ALIGNMENT [1]

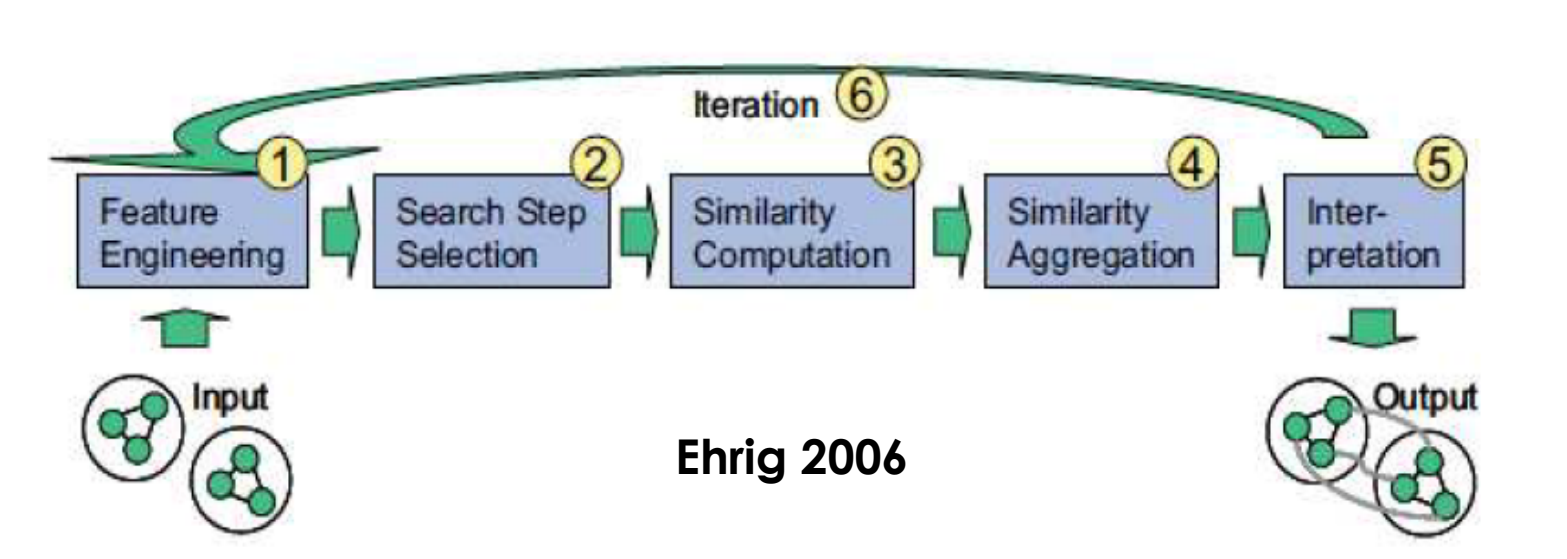


(Lambrix, P., Tan, H.., Liu, Q. 2006)

OA systems

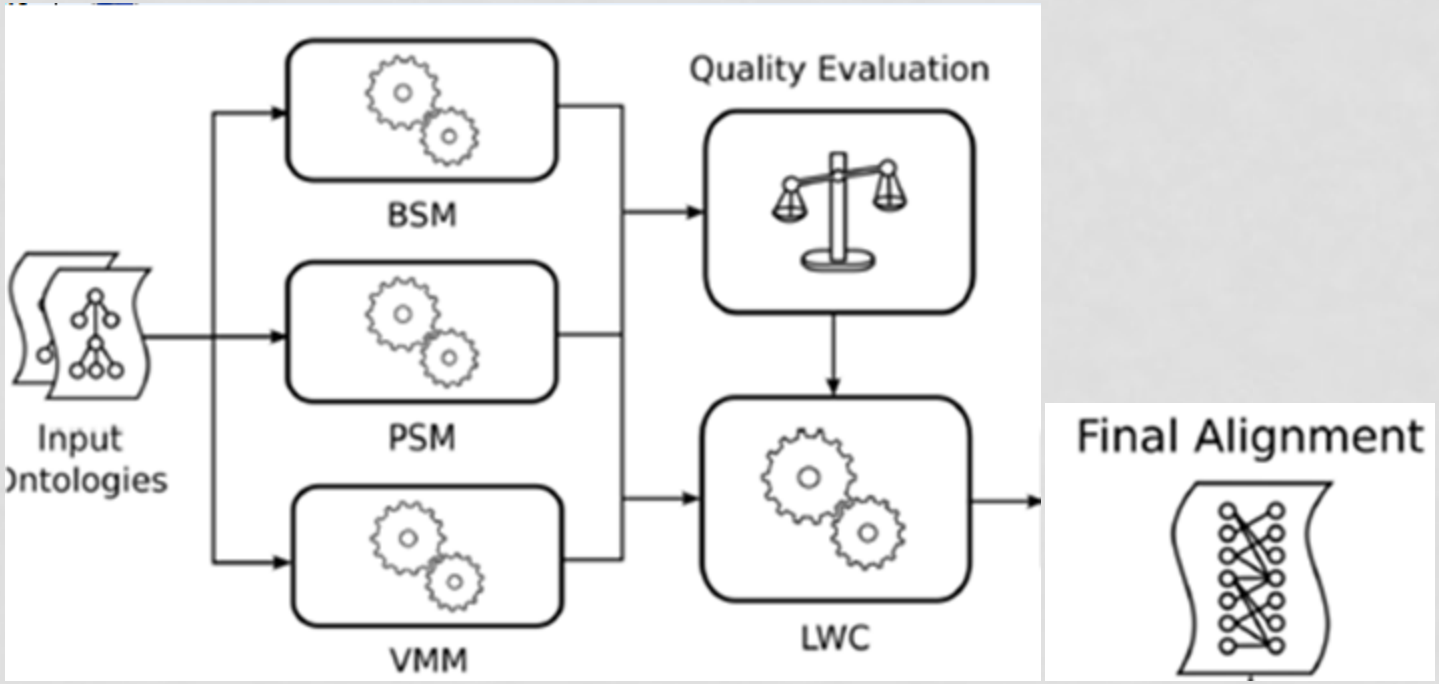
- **take as input a source ontology O_S and a target ontology O_T**
- **typically produces a set M_{ST} of mapping pairs (s_i, ti) between**
- **each pair has a similarity degree d_{sim} in (0, 1].**
- **mapping indicates concept s_i in O_S is similar to the concept t_i in O_T with d_{sim}.**

GENERAL ARCHITECTURE OF OA [2]



1. Select features describing entities.
2. Chose entity pairs to compare.
3. Calculate similarity for entity pairs.
4. Aggregate similarity values for a single pair
5. Decision of alignment
6. New alignment found can influence other pair's similarity value.

A CONFIGURABLE ARCHITECTURE FOR AGREEMENTMAKER [3]



BSM: Base Similarity Matcher

PSM: Parametric String-based Matcher (substring + edit-dist)

VMM: Vector-based Multi-term Matcher (TF-IDF + Cosine sim)

LWC: Linear Weighted Combination (local-confidence weighting scheme)

SEMANTIC SIMILARITY MEASURES

Assess the degree of similarity between two concepts of the SAME ontology.

Path-based :

- Relies on the distance between two entities in the ontology

Information content based :

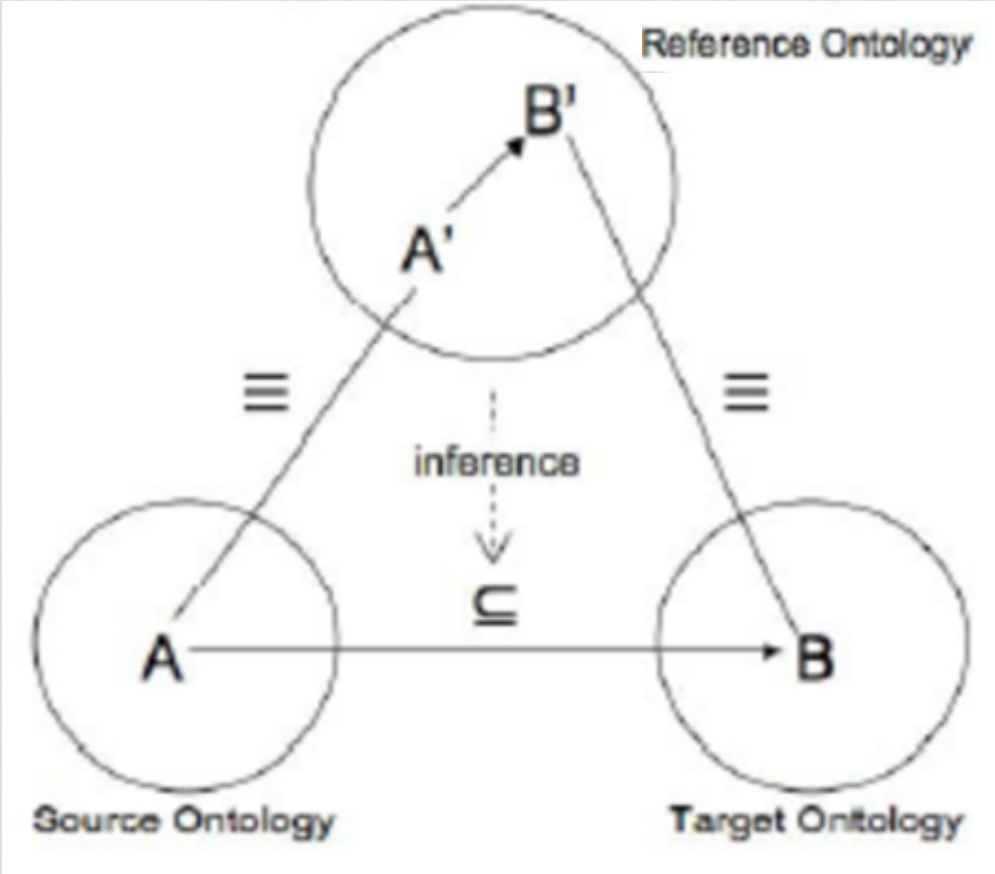
- Information content shows how specific or general a concept is either within the ontology or with respect to a corpus

Set-based:

- Use set operations on sets of features descrbing the two concepts.

Semantic similarity measures have been added to AgreementMaker software based on Xueheng Hu's masters thesis research

USING A REFERENCE ONTOLOGY



(Sabou, M., Aquin, M., and Motta, M. 2007)

RECENT USE OF REFERENCE ONTOLOGIES

Two very recent experiments using reference ontologies to improve the alignment mapping process.

- **ICBO 2011 (Gross *et al*), *Mapping Composition for Matching Large Life Science Ontologies (called Intermediate Ontology)***

- ***OM Workshop 2011 (Cruz et al) Using AgreementMaker to Align Ontologies for OAEI 2011 (called Mediating Ontology)***

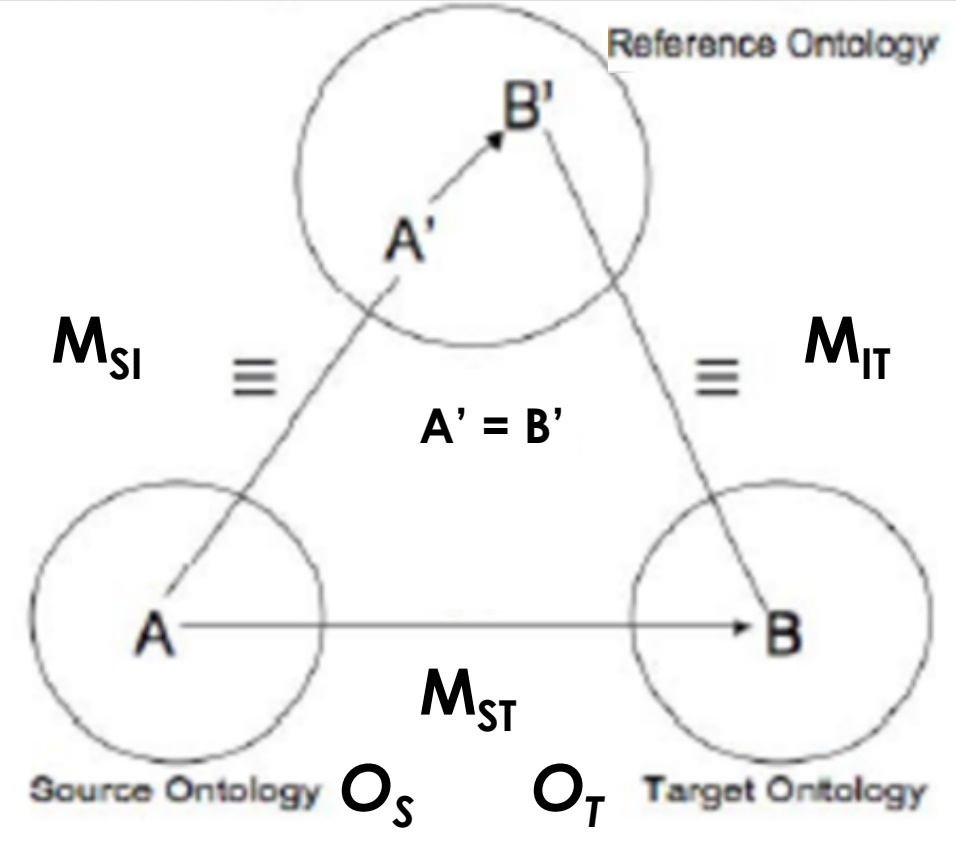
Differences exist in

1) alignment methods for mapping O_S and O_T to O_R

2) aggregation method of similarity values to produce final mapping M_{ST}

Neither incorporates semantic similarity measures between concepts within the reference ontology

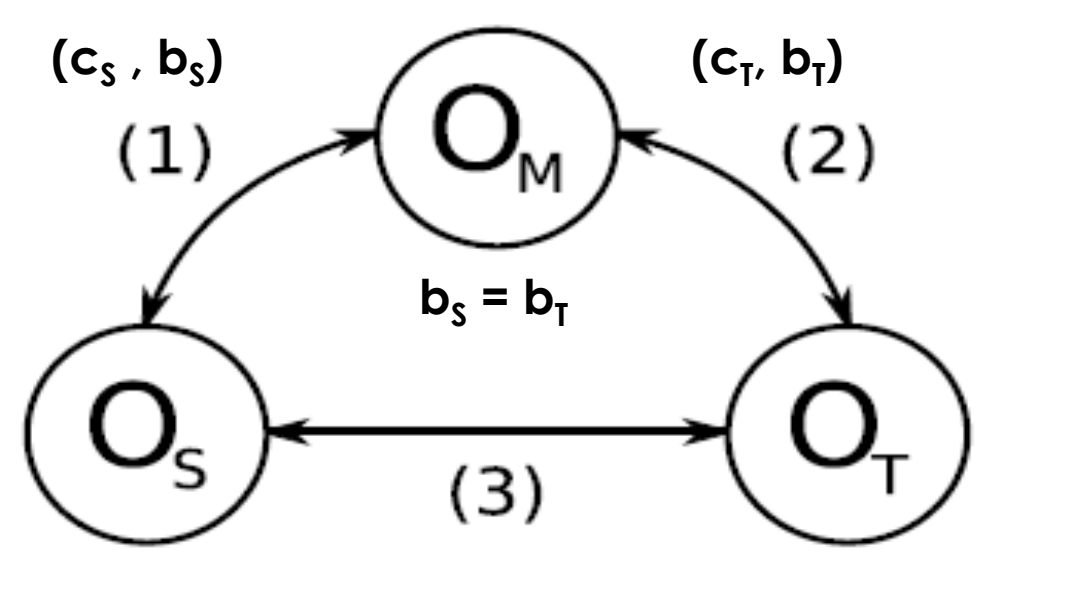
COMPOSITION-BASED MATCHING



- aggSim combines the mapping similarities for M_{SI} and M_{IT}.
  - Average was used.
- Suggest M_{SI} and M_{IT} could be existing mappings as in BioPortal.
  - In experiments determined using linguistic trigram similarity between concept names and synonyms with threshold of 0.8.

$$M_{ST} = \{(c_S, c_T, aggSim(mapSim_{SI}, mapSim_{IT})) \mid c_S \in O_S, c_I \in O_I, c_T \in O_T : \mid \exists(c_S, c_I, mapSim_{SI},) \in M_{SI} \wedge \exists(c_I, c_T, mapSim_{IT},) \in M_{IT}\}$$

AGREEMENTMAKER MEDIATING MATCHER (MM)

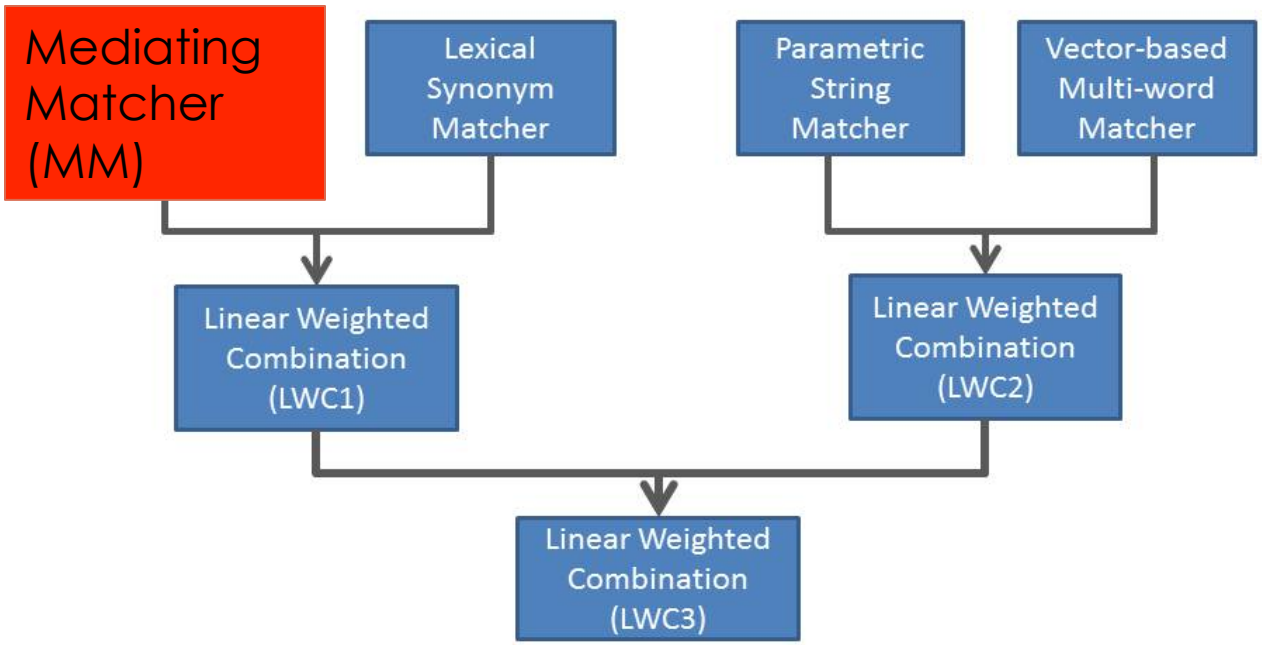


Steps 1 and 2 : Match source and target to mediating ontology using BSM (lexical matcher)

Step 3: source concept is aligned to target concept if (c_S, b_S) and (c_T, b_T) are bridge mappings and $b_S = b_T$

aggSim is the product, $sim(c_S, c_T) = sim(c_S, b_S) * (c_T, b_T)$

OAEI 2011 AGREEMENT MAKER MATCHER STACK (ANATOMY TRACK)



Combination of Matchers in OAEI2011 Matcher

AgreementMaker had the best OAEI 2011 performance with respect to F-measure (91.7%). It used only Uberon and its results are better than those in (Gross et al., 2011)even with its four intermediate ontologies

MEDIATING MATCHER WITH SEMANTIC SIMILARITY *(MMSS)*

- Recent approaches rely on source and target concepts mapping to the *same* concept in the reference ontology
- Using semantic similarity in reference ontologies to find indirect mappings where source and target concepts map to different concepts in mediating ontology
- Semantic similarity is high usually in siblings or parent/ child concepts.

MEDIATING MATCHER WITH SEMANTIC SIMILARITY *(MMSS)*

Mapping set M_{ST} consisting of cases where c_s maps to b_s and c_t maps to b_t and $b_s=b_t$

U_S – set of unmapped source concepts
U_T –set of unmapped target concepts

For each (s,t) in U_S X U_T consider all bridge mappings tied to s and t.

Add max of (*s, t, agg(mapSim$_{SI}$, mapSim$_{TI}$, bridgeSim))* to E_{ST} subject to threshold value

Final mapping is M_{ST} U E_{ST}

ONTOLOGY ALIGNMENT EVALUATION INITIATIVE (OAEI)

- Annual event for evaluation of ontology alignment methods
- Coordinated international effort to evaluate, compare and improve alignment approaches
- Participation of several OA systems on several "tracks" that are specialized for different purposes
- OAEI2011 consisted of benchmark, anatomy, conference and instance matching tracks
- Research focus so far on anatomy track since it involves two real-world ontologies which produce considerably more mappings than other tracks

INITIAL EXPERIMENTS

- Using the OAEI anatomy track
  - Mouse Anatomy & Human Anatomy ontologies
- Mediating Ontology used:
  - Uberon
- Semantic Similarity used :
  - Lin (IC based)
- Aggregator used:
  - minimum

PERFORMANCE MEASURES

- Given a reference alignment (R) and a produced alignment (A), the following measures are calculated for evaluation
- *Precision:* measures the percentage of output alignments that are actually correct.
  - $P(A, R) = |R \cap A| \ / \ |A|$
- *Recall:* measures the percentage of correct alignments that are output by the OA system
  - $R(A, R) = |R \cap A| \ / \ |R|$
- *F-measure:* combination of precision and recall with emphasis on one or both
  - $F(A, R) = (b^2 + 1) \cdot P(A, R) \cdot R(A, R) \ / \ b^2 \cdot P(A, R) + R(A, R)$

TWO COMPARISONS

1) Only the mappings from the MM are compared to only those from the MMSS with varying semantic similarity thresholds

2) OAEI 2011 AgreementMaker final mappings using MM compared to final mappings replacing MM with MMSS to investigate the effects of its LWC matchers combining the various mappings results

INITIAL RESULTS

	Mapped	Correct	Precision	Recall	F-measure
MM	1200	1143	95.2	75.4	84.2
MMSS, 0.0	1322	1152	87.1	76	81.2
MMSS, 0.65	1301	1151	88.5	75.9	81.7
MMSS, 0.85	1240	1150	92.7	75.9	83.5
MMSS, 0.90	1229	1148	93.4	75.7	83.6
OAEI 2011					
MM	1443	1350	93.6	89.1	91.2
MMSS, 0.85	1447	1348	93.2	88.9	91.0
MMSS, 0.90	1447	1350	93.3	89.1	91.1

DISCUSSION

- Unmodified Mediating Matcher (more precision) vs modified Mediating Matcher (more recall)
- OAEI2011 matcher using modified mediating matcher has same recall as OAEI2011 matcher using unmodified mediating matcher
- Point of difference : Three new correct mappings found, at the same time three correct mappings were lost

LOST 3 AND NEW 3 MAPPINGS

MA Source	HA MMSS Target	Uberon B_S	Uberon B_T
gastrointestinal system mesentery	Mesentery	gastrointestinal system mesentery	Mesentery
Limb long bone	Long bone	Limb long bone	Long bone
Brain ependyma	Ependyma	Brain ependyma	Ependyma

New Mappings, OAEI MMSS but not OAEI MM

VSM found, PSM did not

PSM found, VSM overrided it

No other matcher

MA Source	HA MM Target	Uberon B_S	Uberon B_T
Brain arach-noid matter	Cerebral Arachnoid Membrane	Brain arach-noid mater	leptomeninges
Iliac circum-flex artery	circumflex iliac artery	Iliac circum-flex artery	Deep circumflex iliac artery
Vagina squamous epithelium	Vagina squamous epithelium	Vagina squamous epithelium	Stratified squamous epithelium

Lost Mappings, OAEI MM but not OAEI MMSS

PSM found
VSM found

PSM found
VSM found

PSM found

FUTURE RESEARCH

- Try to preserve the 3 lost mappings while gaining 3 new mappings by studying interaction between various LWC matchers, possibly adding own quality heuristics
- Experiment with different semantic similarity measures and aggregations
- Using a other reference ontologies besides Uberon for the anatomy track
- Combination of different reference ontologies.

THANK YOU

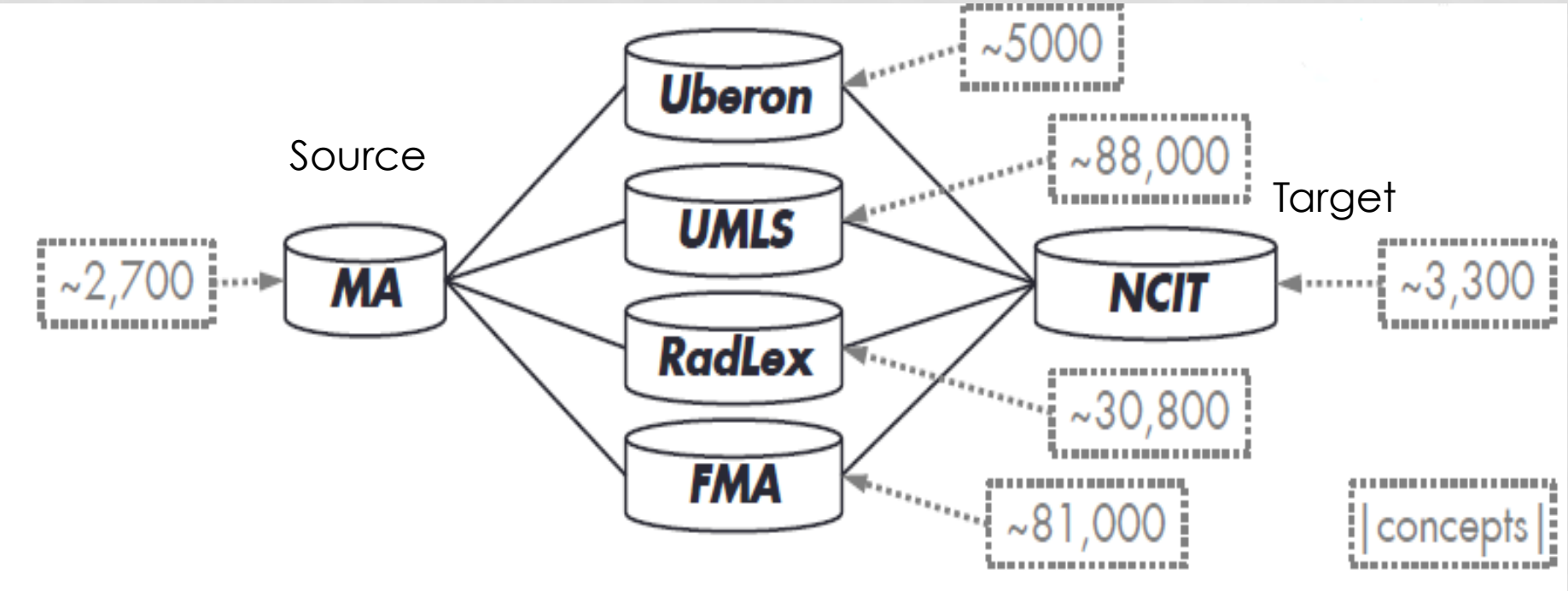
QUESTIONS?

EXISTING OA SYSTEMS USING SEMANTIC SIMILARITY

System	Approach	Semantic Similarity used
OLA (Euzenat and Valtchev 2003)	Pair of identifiers broken into sets of atomic terms. SS between pairs of terms from the set calculated	Wu-Palmer
Imapper (Su 2004)	Similarity of mapping can be strengthened through SS from Wordnet using descriptive labels of concepts	Simple path-based measures
ASMOV (Jean-Mary et al. 2009)	SS is used when the concepts don't match exactly and are not in same Wordnet synset	Lin
CIDER (Gracia and Mena 2008)	SS calculated based on different senses and synonyms of keyword	Sense similarity
UFOme (Pirro and Talia 2010)	SS between synsets of concept terms, again uses Wordnet	Lin

INTERMEDIATE ONTOLOGIES

Intermediate Ontologies



- ***(Gross et al), Mapping Composition for Matching Large Life Science Ontologies***