Ontology Mapping and Alignment

Natasha Noy
Stanford University

Outline

- *Semantic integration*
- *Types of ontology mismatches*
- *Mapping discovery*
- *Using mappings*
- *Challenges/Issues*

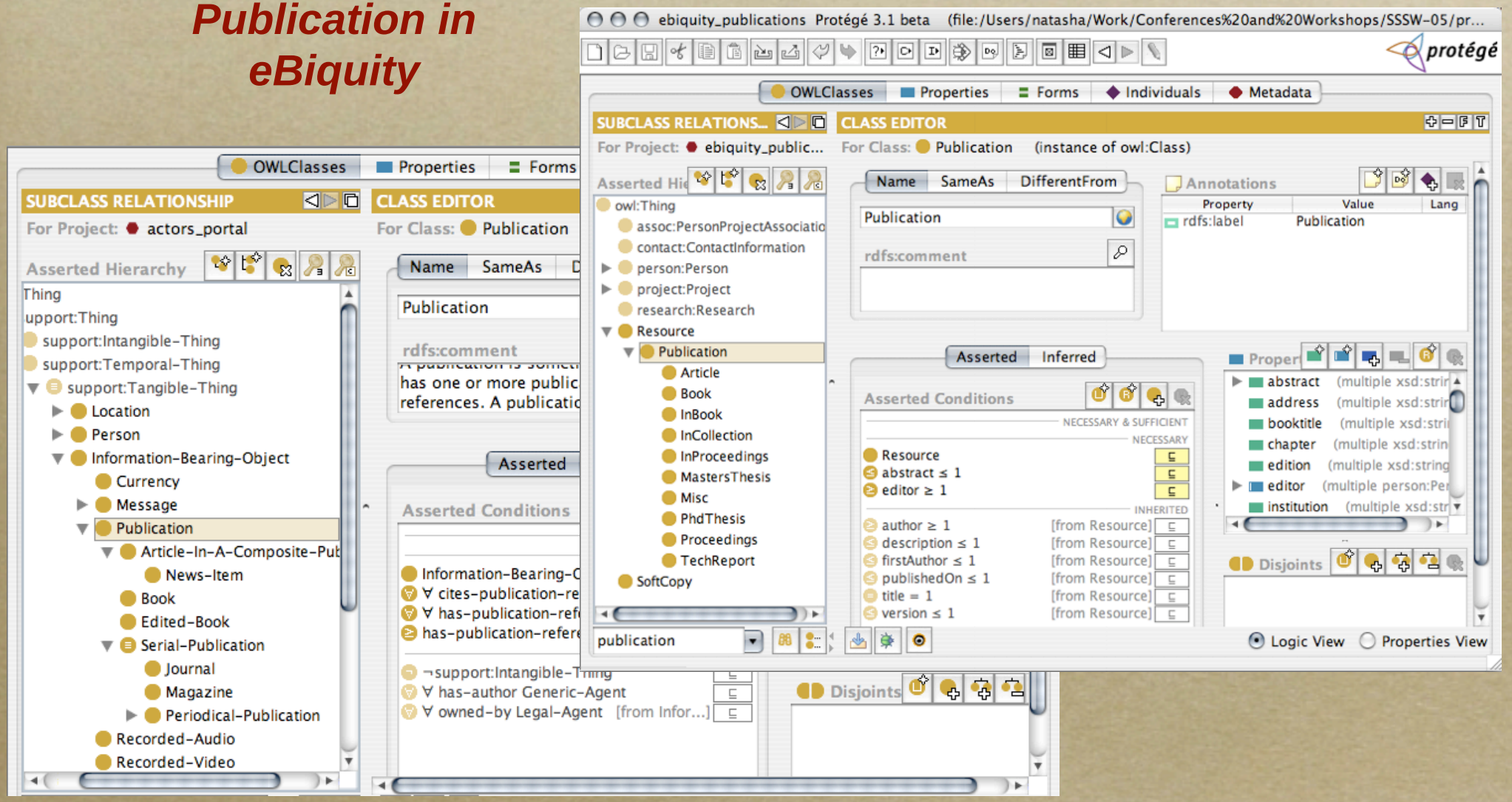
Lots of Overlapping Ontologies on the Semantic Web

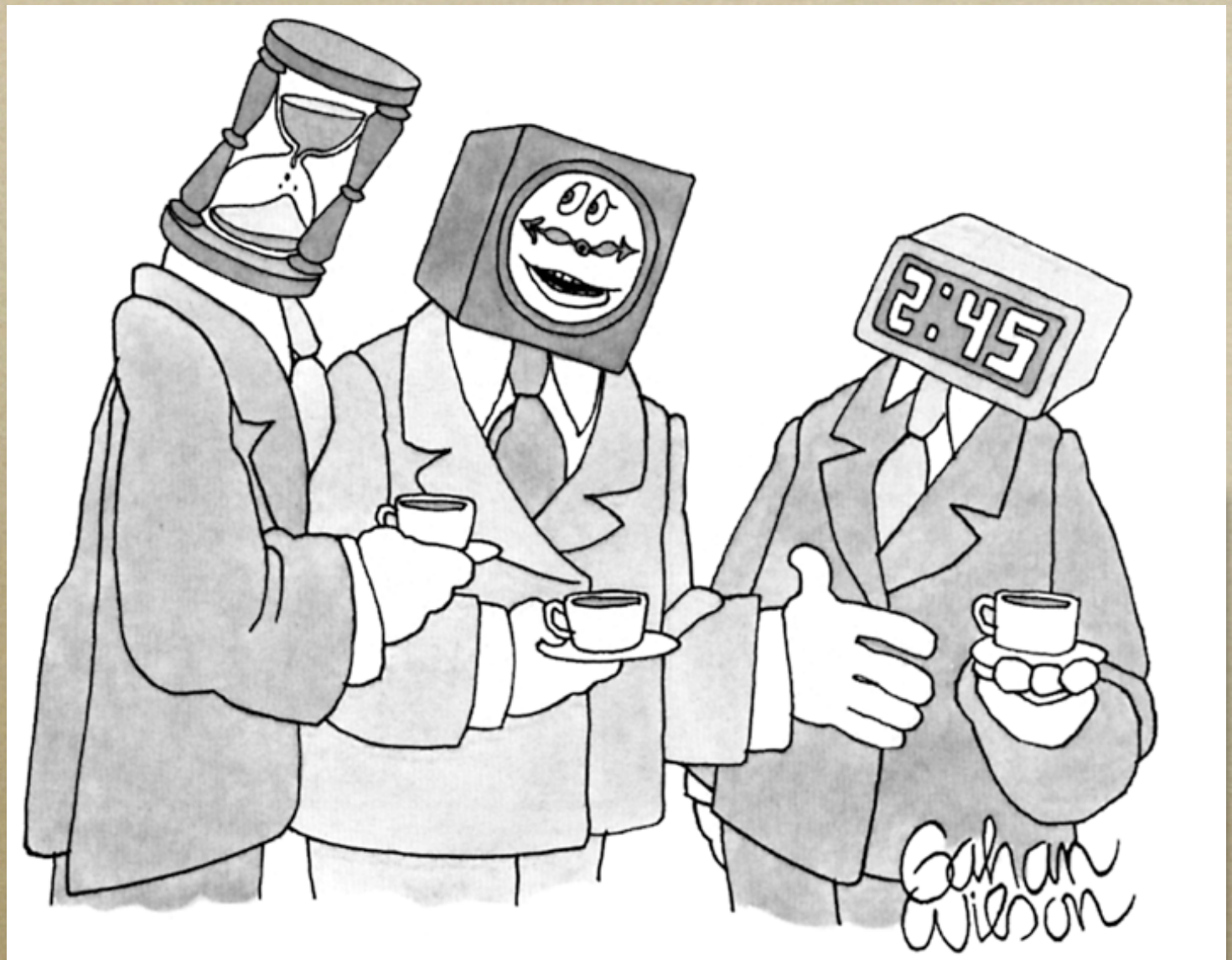
- *Search Swoogle for "publication"*
- *185 matches in the repository*
- *Different definitions, viewpoints, notions*

Example Definitions of School

Publication in AKTors

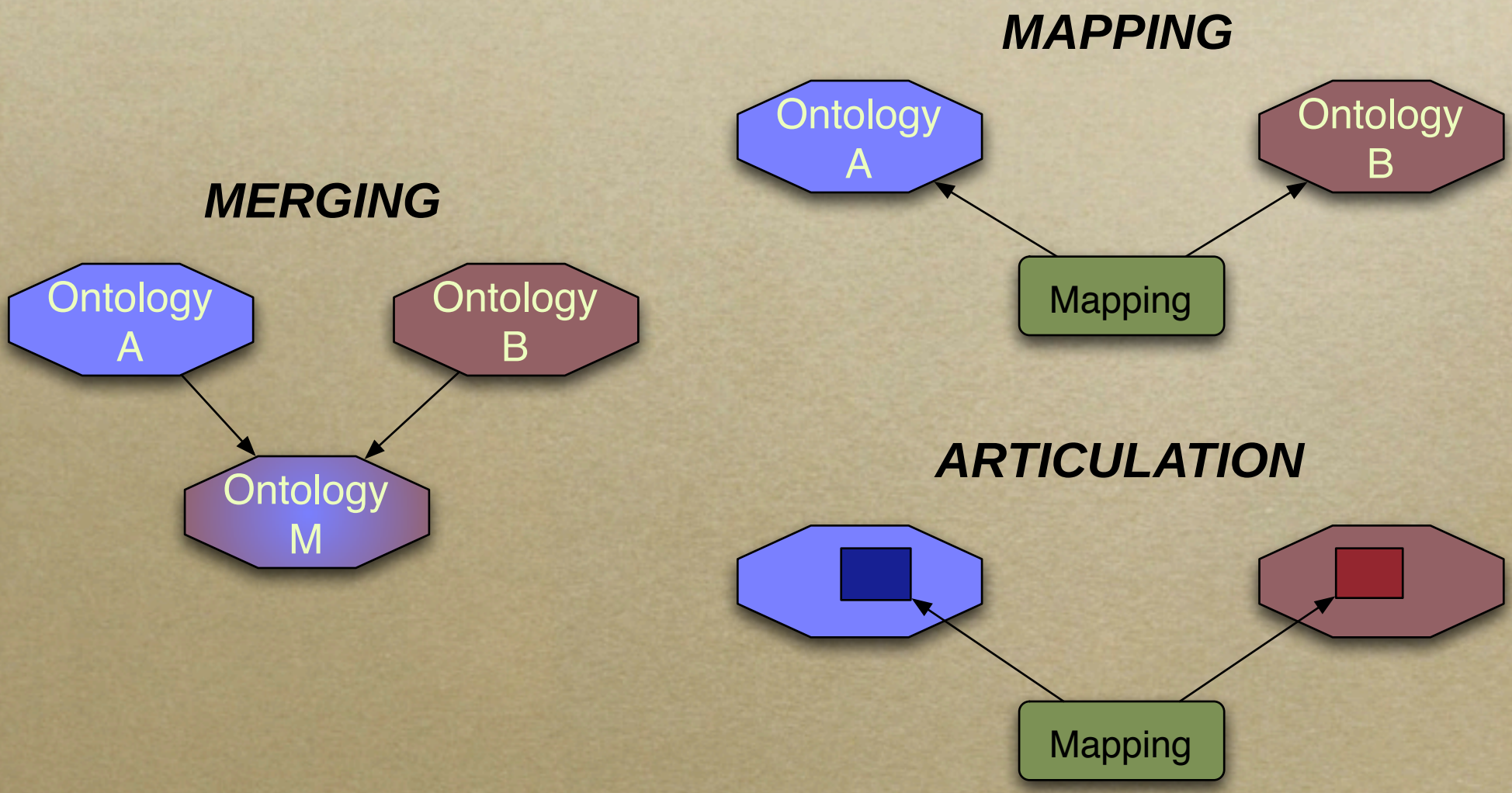
Publication in eBiquity





"Basically, we're all trying to say the same thing."

Creating Correspondences Between Ontologies



Semantic Integration Tasks

- *Queries across multiple resources*
- *Data transformation*
- *Reasoning with mappings*

Outline

- *Semantic Integration components and tasks*
  - *Types of ontology mismatches*
- *Mapping discovery*
- *Using mappings*
- *Challenges/Issues*

Reasons for Mismatches

- *Ontology is not a reality it is a subjective representation of it*
  - *Different designers have different views*
- *Different tasks and requirements for applications*
- *Different conventions, etc.*

Types of Mismatches

- *Language-level mismatches*
  - *Difference in expressiveness or semantics of ontology language*
- *Ontology-level mismatches*
  - *Difference in the structure of semantics of the ontology*

Language-level Mismatches

- *Syntax*
- *Expressiveness*
  - *e.g., presence of disjoints, negations, expressions, unions, intersections, metaclasses, etc. in the language*
- *Semantics of primitives*
  - *e.g., union vs intersection semantics for multiple domain and range declarations*

Ontology-level Mismatches

- *The same terms describing different concepts*
- *Different terms describing the same concept*
- *Different modeling paradigms*
  - *e.g., intervals or points to describe temporal aspects*
- *Different modeling conventions*
- *Different levels of granularity*
- *Different coverage*
- *Different points of view*
- *…*

Ontology-level Mismatches: Examples

- *Compare ontologies*
  - `http://www.aktors.org/ontology/`: *the ontology used in CS AKTive Portal testbed*
  - `http://ebiquity.umbc.edu/ontology/`: *the ontology developed by the UMBC eBiquity group*

Some of the Differences

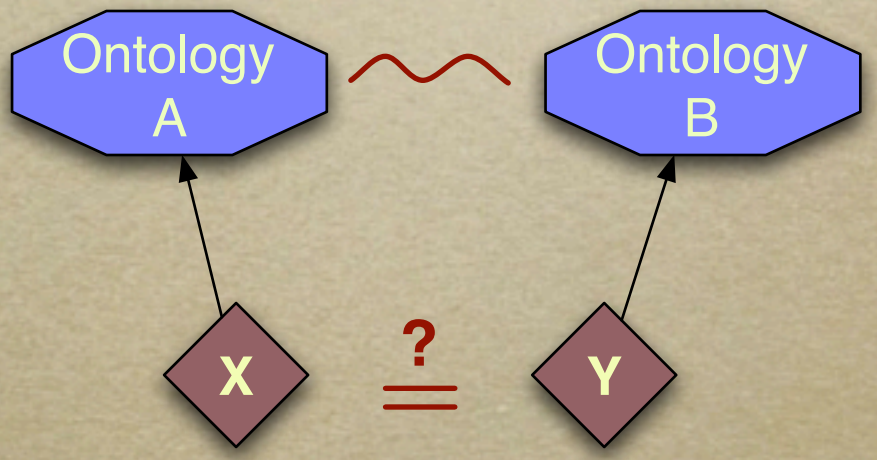
Different names for the same concept	`PhD-Student`	`PhDStudent`
Same term for different concepts	`Project`: Only current projects	`Project`: Past projects and proposals
Scope	Includes periodicals, composite publications...	Includes alumni, guest speakers, etc
Different focus in definition	`Publication`: Restrictions on citations, refs.	`Publication`: Restrictions on abstract, editor
Constructs used	Includes defined classes	Only primitive classes
Different modeling conventins	`Journal` is a class	`journal` is a property
Granularity	`Professor-In-Academia`	Adjunct, affiliated, associate, principal, etc.
Different modeling conventions and level of detail	`address` property broken up into several properties	`address` property is a single string property

Outline

- *Semantic Integration components and tasks*
- *Types of ontology mismatches*
- *Mapping discovery*
- *Using mappings*
- *Challenges/Issues*

Categories of Mappings

- *Ontology-to-ontology mapping*
- *Data matching*



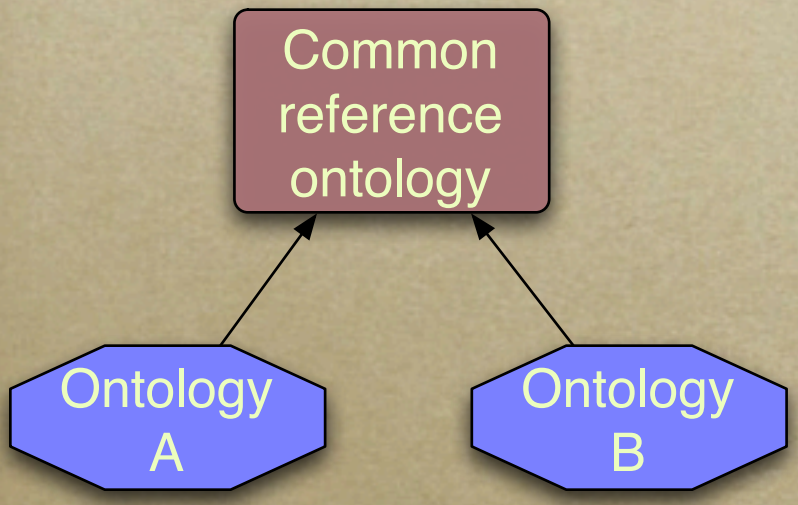
Mapping Discovery

- *Information Sources*
- *Methods*

Mapping Discovery: Information Sources

- *A common reference ontology*
- *Lexical information*
- *Ontology structure*
- *User input*
- *External resources*
- *Prior matches*

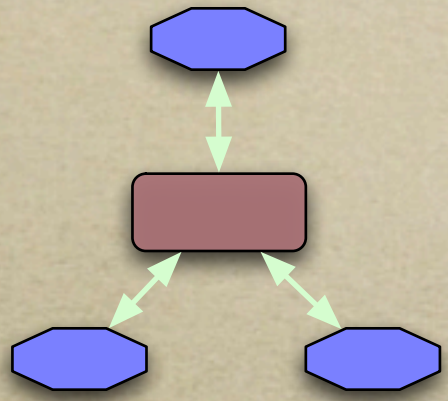
Using a Common Reference Ontology

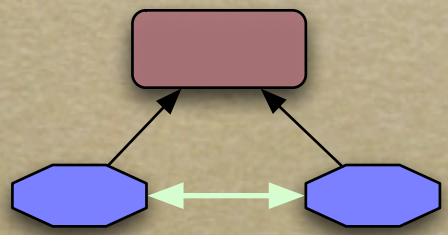


- *"Upper" ontologies designed to support information integration*
  - *designed in principled way*
  - *provide common reference terminology*
  - *SUMO, DOLCE*
- *Domain-specific interlingua*
  - *PSL*

Solve the problem before it arises

- *Modes of mapping*
  - *Ontology-to-interlingua*
  - *Ontology-to-ontology, using the interlingua information*





Using reference ontologies: Problems

- *People are reluctant to reuse*
- *There have been some successes (in domain-specific settings) and failures*
- *Usual problems with having standards*

Using Lexical Information

- *String normalization*
  - *upper and lower case*
  - *blanks and delimiters*
  - *diacritics*
  - *stop-words*
- *String distance*
  - *Hamming distance*
  - *edit distance*
- *Soundex*
- *Thesaurus*

Using Lexical Information

- *Ed Hovy (USC/ISI)*
  - *provides a set of heuristics for aligning domain ontologies to a central ontology*
  - *uses natural-language analysis of concept names and definitions*
    - *splitting composite names*
    - *finding common substrings*
    - *finding the ratio of common words in definitions*
  - *uses hierarchy information*

Using Ontology Structure

- *IF-Map (Kalfoglou, Schorlemmer)*
- *Using metrics to compare OWL concepts (Euzenat and Volchev)*
- *QOM (Ehrig and Staab)*
- *Similarity Flooding (Melnik, Garcia-Molina, Rahm)*
- *Chimaera (Stanford KSL)*
- *Prompt and AnchorPrompt (Stanford SMI)*
  - *a number of others...*

Using External Sources

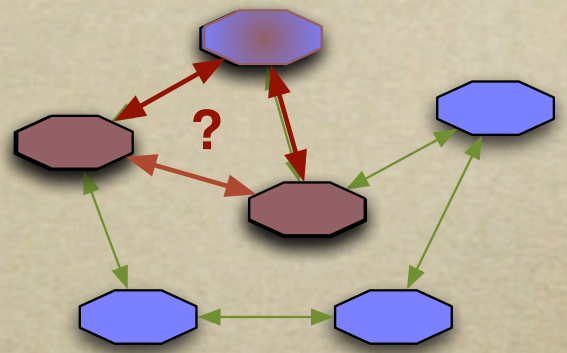
- *Grounding sources in standard terminologies*
  - *WordNet*
  - *UMLS*
  - *S-Match (U. Trento)*

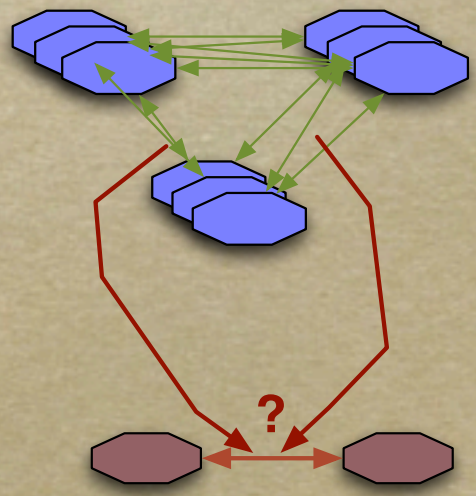
User Input

- *Providing information on initial alignment*
- *Providing feedback on alignments*
- *Invoking or composing alignment methods*
  - *Prompt, Chimaera, ONION*

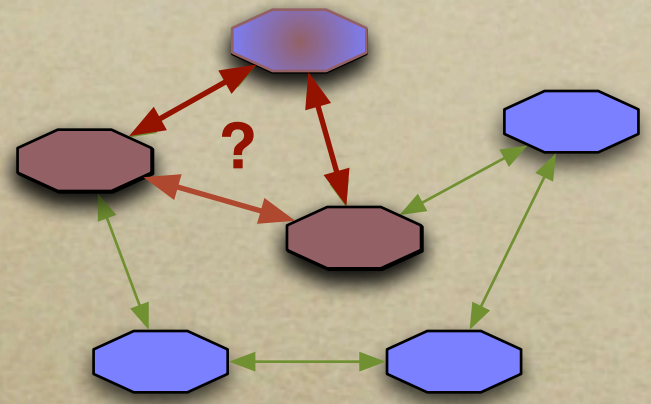
Using Prior Matches

- *Composing existing matches*
  - *Semantic gossiping*
- *Using a corpus of matches*
  - *Alon Halevy (UW)*





Mapping Composition

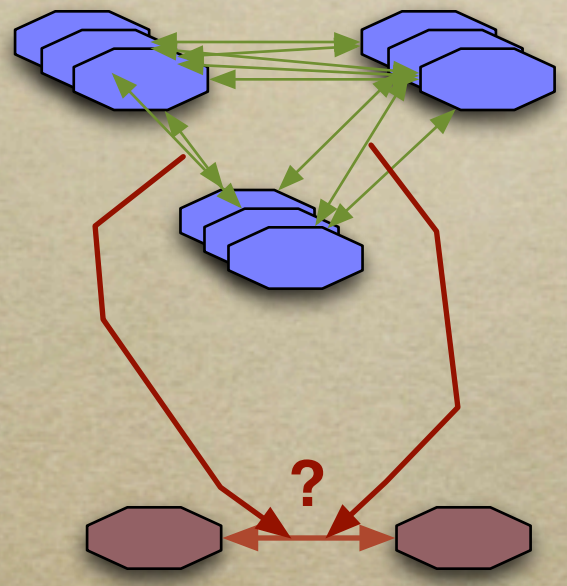


- *Issues*
  - *Quality of initial mappings*
  - *Composition without loss of information*
  - *Choosing which composition route to follow*

Semantic gossiping, Piazza (UW)

Using Corpus of Matches

- *Contents of the corpus*
  - *Domain models*
  - *Instance data*
  - *Validated mappings*
  - *Queries*
  - *Meta-data*
- *Statistics on the corpus*
  - *Term usage*
  - *Co-occurrence of schema and ontology*



O.Etzioni, A. Halevy, et.al. (UW)

Mapping Discovery: Information Sources

- *A common reference ontology*
- *Lexical information*
- *Ontology structure*
- *User input*
- *External resources*
- *Prior matches*

Mapping Methods

- *Heuristic and Rule-based methods*
- *Graph analysis*
- *Machine-learning*
- *Probabilistic approaches*
- *Reasoning, theorem proving*

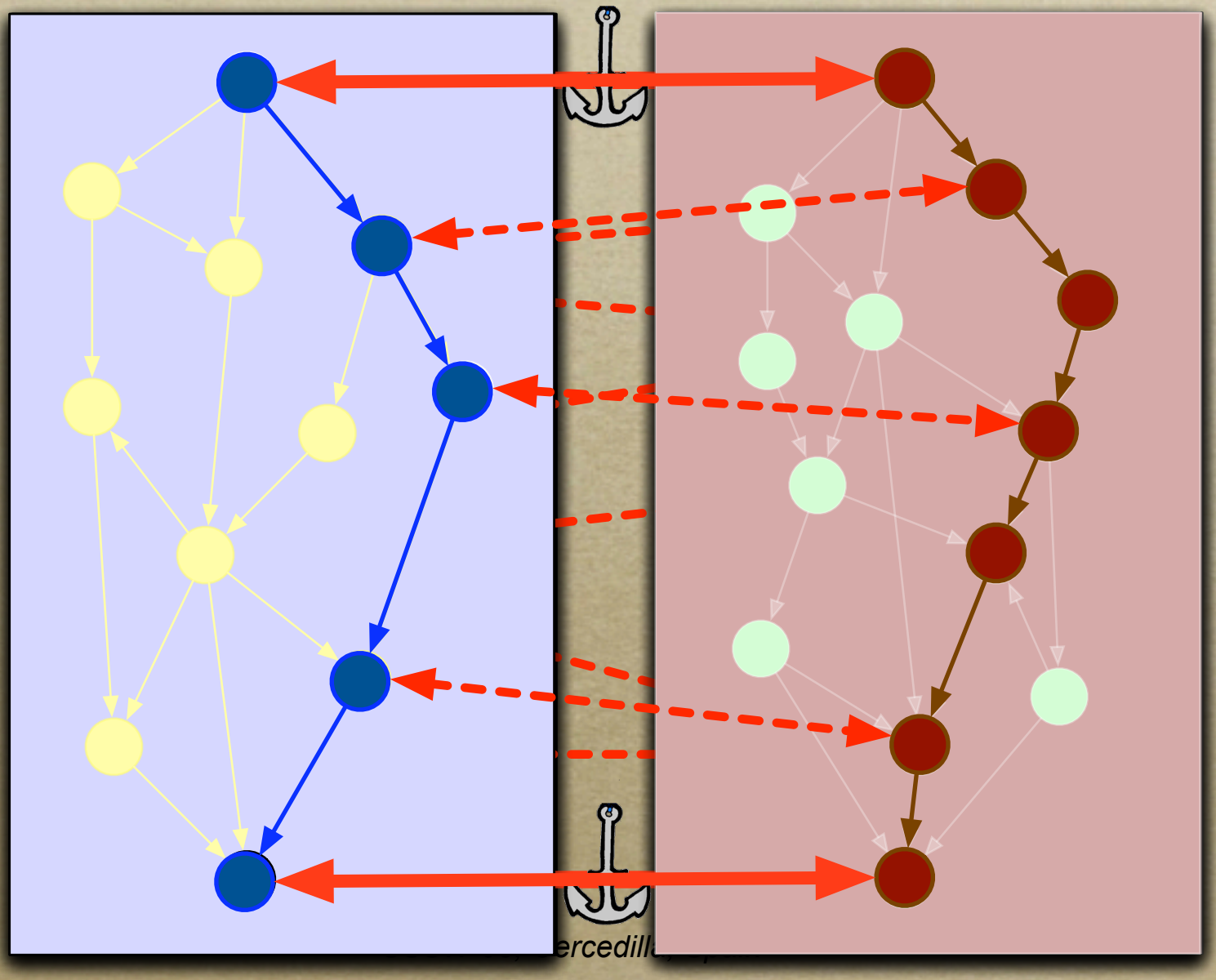
Rule-Based and Graph-Analysis Methods

- *Rule-based methods*
  - *Most structure-analysis and lexical analysis methods*
    - *Prompt, Chimaera, QOM, IF-Map...*

Graph-based Methods

- *Treat ontologies as graphs and compare the corresponding subgraphs*
  - *Similarity Flooding, Anchor-Prompt*

AnchorPrompt: Analyzing Graph Structure

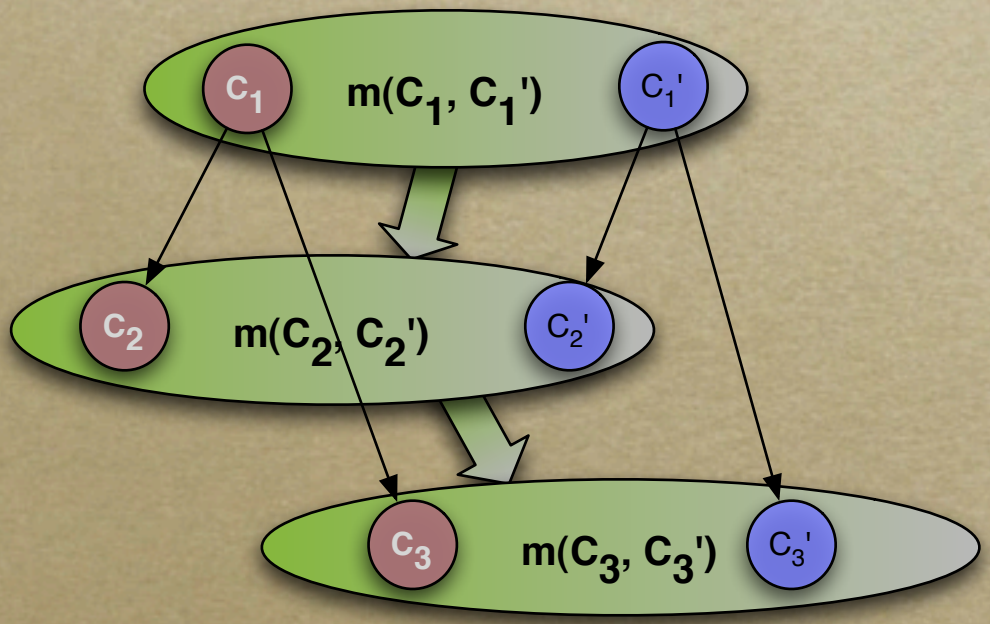


Machine Learning Approaches

- *Statistics of data content*
- *Using multiple learners*
- *Using instance and values information*
  - *GLUE, LSD, SemInt*

Probabilistic Approaches

- *Combining results produced by heuristic-based mappings*
  - *OMEN (Mitra & Noy)*



- *More in data matching*

Reasoning and Theorem Proving

- *We are mostly an AI crowd, after all...*
  - *S-Match*
    - *Start with a combination of matchers using lexical information and external resources*
    - *Use a SAT solver to find equivalence, generalization, and specialization mappings*

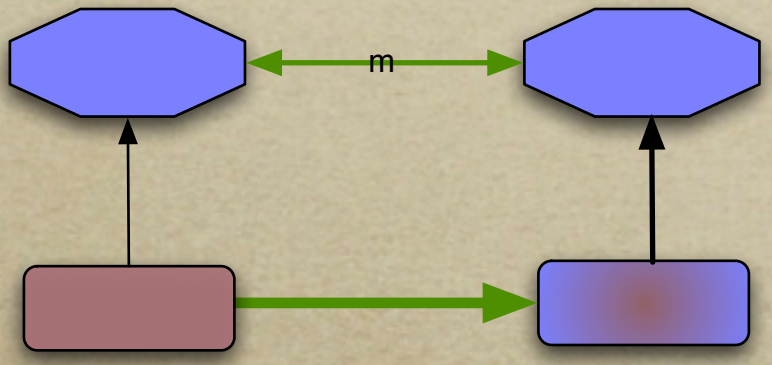
Outline

- *Semantic Integration components and tasks*
- *Types of ontology mismatches*
- *Mapping discovery*
- *Using mappings*
- *Challenges/Issues*

Using Mappings

- *Data transformation*
- *Query answering*
- *Reasoning with mappings*
  - *mapping composition (covered earlier)*
- *Generation of ontology extensions*

Data Transformation

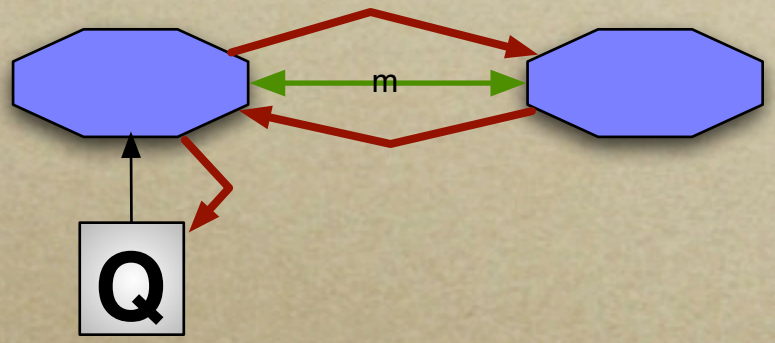


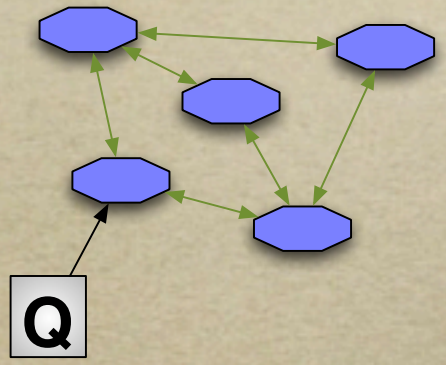
- *Mapping Interpreter (Stanford SMI), OntoMerge*

Data Transformation (II)

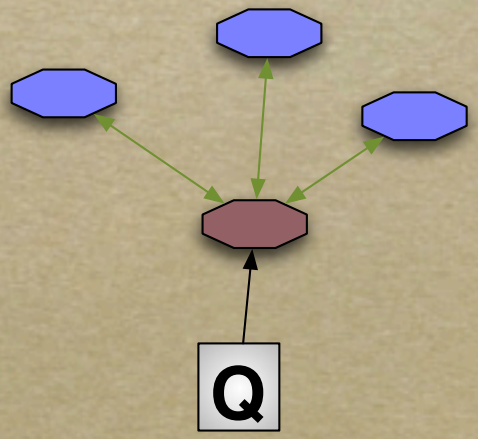
- *Mapping interpreter (Stanford SMI)*
  - *Uses an instantiated ontology of mappings*
    - *mapping structure*
    - *Python rules*
- *OntoMerge*
  - *Treats source ontologies with data and mapping axioms as a single ontology*
  - *Uses a theorem prover to create new data*

Query Answering

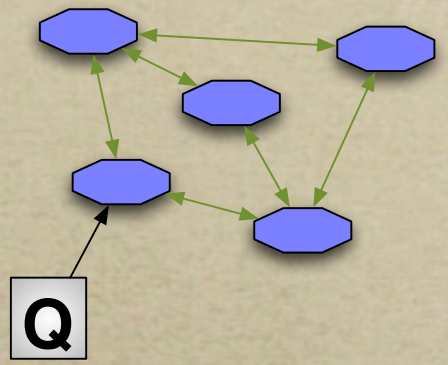




- *Two settings*
  - *one-to-one mappings*
  - *global ontology*

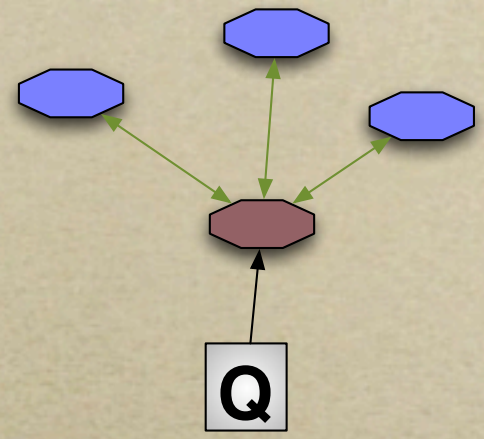


Query Answering (II)



- *Piazza (UW)*
  - *Peer-to-peer architecture for query answering*
  - *Query refomulation using mappings between adjacent peers*

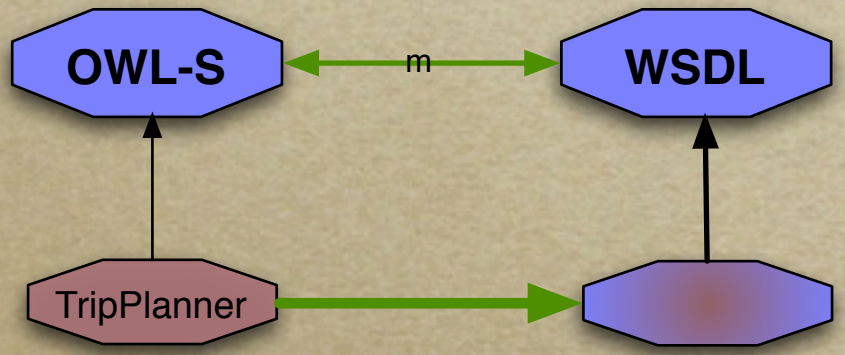
Query Answering



- *OIS (Calvanese, et. al.)*
  - *Global ontology mapped to local ontologies*
  - *Mappings defined as views*
  - *Using a Description Logic Reasoner to answer queries*

Generation of Ontology Extensions

- *OntoMerge*



Outline

- *Semantic Integration components and tasks*
- *Types of ontology mismatches*
- *Mapping*
- *Using mappings*
- *Challenges/Issues*

Challenges/Issues

- *Design space of mapping approaches*
  - *Can we create a "toolbox" for designing mapping approaches that fit a given problem?*
  - *We have identified some components, but how can we bring them together?*
- *Have we reached a "ceiling" in mapping discovery?*
  - *Will it be "lots of work for little gain" from now on?*
  - *Are there serious untapped resources?*

Challenges/Issues

- *Are imperfect and inconsistent mappings useful?*
- *How do we maintain mappings when ontologies evolve?*
- *How do we evaluate and compare different tools?*
  - *EON experiment*
  - *NIST experiment*