Semi-Automatic Information and Knowledge Systems
:
Ontology Merging & Integration

Monika Lanzenberger

Outline

- Ontology Reuse
- Integration
- Merging
- Tools

An ontology is a tuple:

$$O := (C, H_C, R_C, H_R, I, R_I, A)$$

Combining ontologies O_1 and O_2 is done by:

Merging Integration	Ontology Reuse
Mapping Alignment	Ontology Matching

[Ehrig, et al. 2004, Kalfoglou and Schorlemmer 2003]

Ontology are artifacts shared by different applications

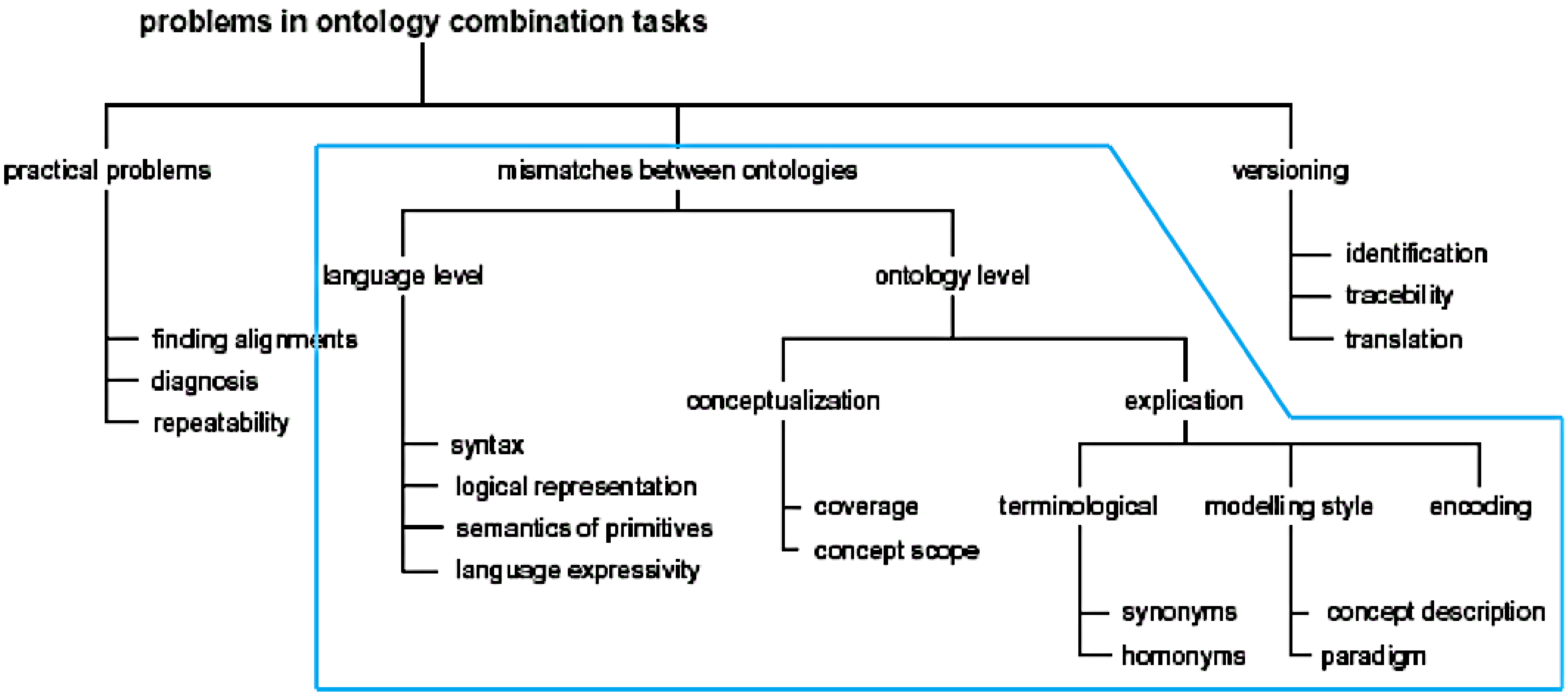
- Create common ontologies
- Extend them for specific domains and applications
- Using same top-level ontology alleviates reuse problems

However, combination problems need to be solved ...

Typical problems when combining ontologies:

- Practical Problems
- Mismatches between ontologies (or entities)
- Synchronization of the changes
  made to source ontologies (Versioning)

[Klein 2001]



[Klein 2001, Su 2004]

Syntax (e.g., Class definition):

- <rdfs:Class ID = "Car"> (RDF Schema)
- (defconcept Car) (LOOM)

Logical Representation (e.g., representing disjointness):

- disjoint A B
- A subclass of (NOT B),B subclass-of (Not A)

Semantics of primitives e.g., same name but different meanings:

- several interpretations of A equalTo B

Expressivity: which notions can be expressed (e.g., negation, quantification etc.)

Conceptualization mismatches:
difference in the way a domain is interpreted (conceptualized), which results in different ontological concepts or different relations between those concepts.

- Model coverage and granularity:
  mismatch in the part of the domain that is covered by the ontology, or the level of detail to which that domain is modeled e.g. , one vehicle ontology might model cars but not trucks
- Scope:
  two classes seem to represent the same concept, e.g. employee is described differently depending on the scope

Ontology (Model) Level Mismatches 2

Explication: difference in the way the conceptualization is specified.

Terminological

- Synonym terms: same thing is represented by different names
- Homonym terms: same term has different meanings depending on the context

Modeling style

- Paradigm:
  Different paradigms can be used to represent time, action, plans, causality, etc.
  E.g., time interval versus point
- Concept description:
  several choices can be made for the modeling of concepts, e.g.,
    dissertation < book < scientific publication < publication
    dissertation < scientific book < book < publication
    or as subclass of both book and scientific publication

Encoding, different formats and different languages

- date dd/mm/yyyy or mm-dd-yy
- miles or kilometers
- Deutsch or English

[Klein 2001, Su 2004]

Two different types of ontology reuse...

merging:

- building an ontology in one subject reusing two or more different ontologies on that subject
- sources are unified into a single one
- it can be difficult to identify unchanged regions
- truly different ontologies, not simple revisions, improvements or variations of the same ontology

integration:

- building an ontology in one subject reusing one or more ontologies in different subjects
- source ontologies are aggregated, combined, assembled together
- possibly after reused ontologies were changed (extension, specialization or adaption)
- identification of unchanged regions taken from source ontologies is easier
- integration is more complex than merging

[Pinto, et al. 1999]

- Ontology Reuse
- Integration
- Merging
- Tools

Integration process takes place along the entire ontology building life cycle

METHONTOLOGY

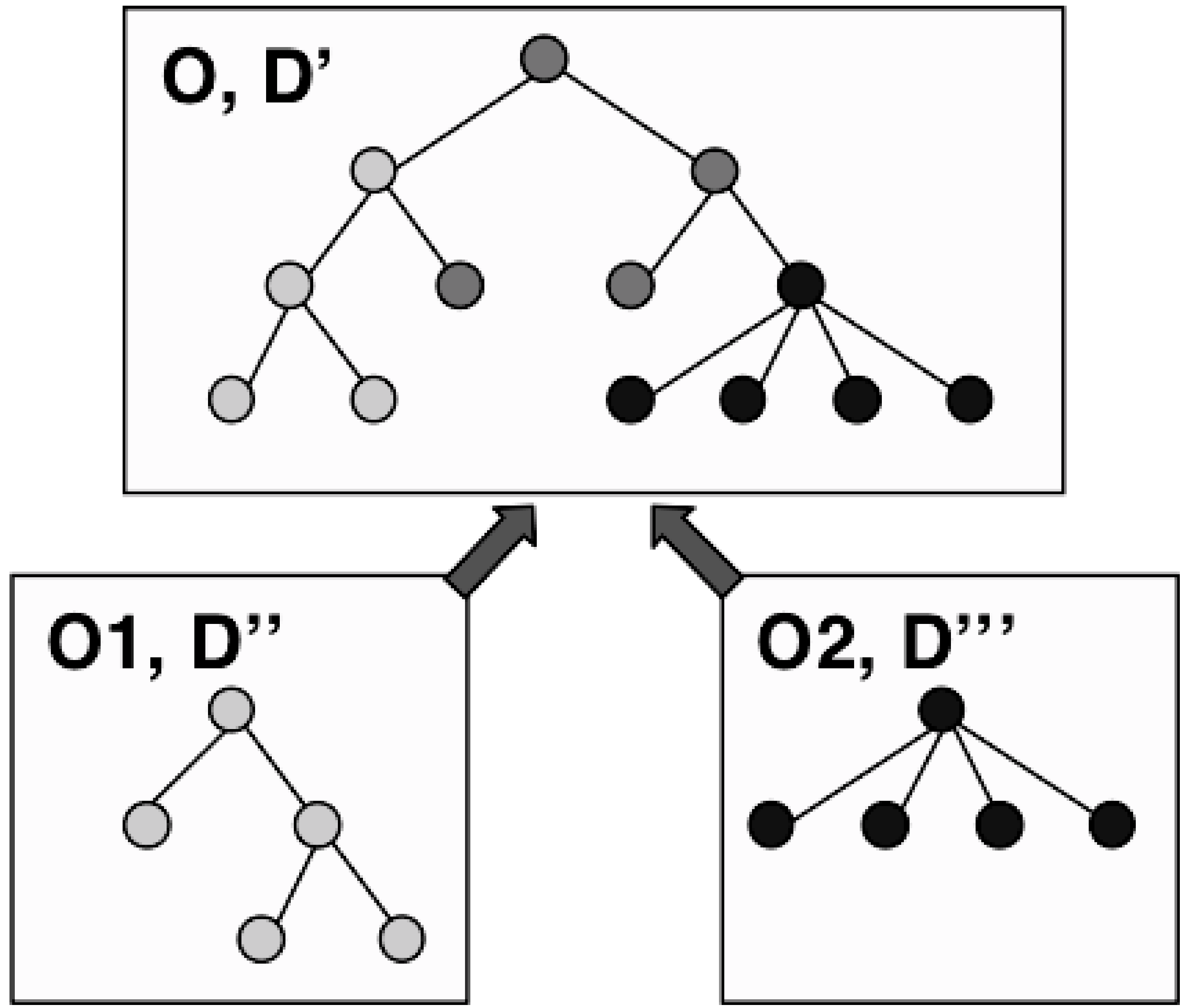
Stages of the building process:

- Specification
- Conceptualization
- Formalization
- Implementation
- Maintenance

... integration should begin as early as possible.

[Pinto and Martins, 2000; Fernández et al. 1999]

Different ontologies:
O, O_1, and O_2
Different domains:
D', D'', and D''



[Jakoniene 2003]

Identify

... the possibility of integration

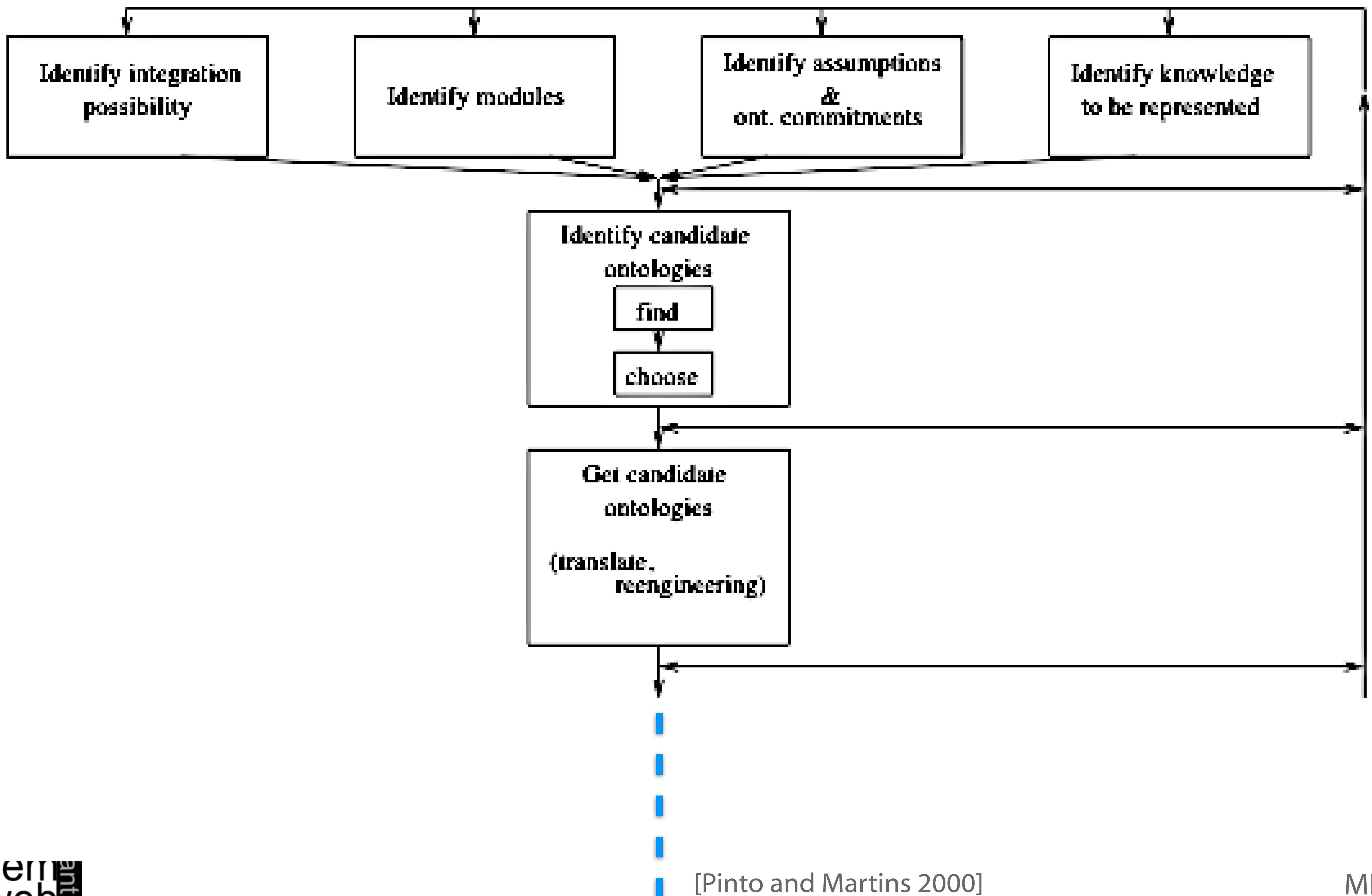
... the modules of the ontology

... the assumptions and ontological commitments for the ontology and each module

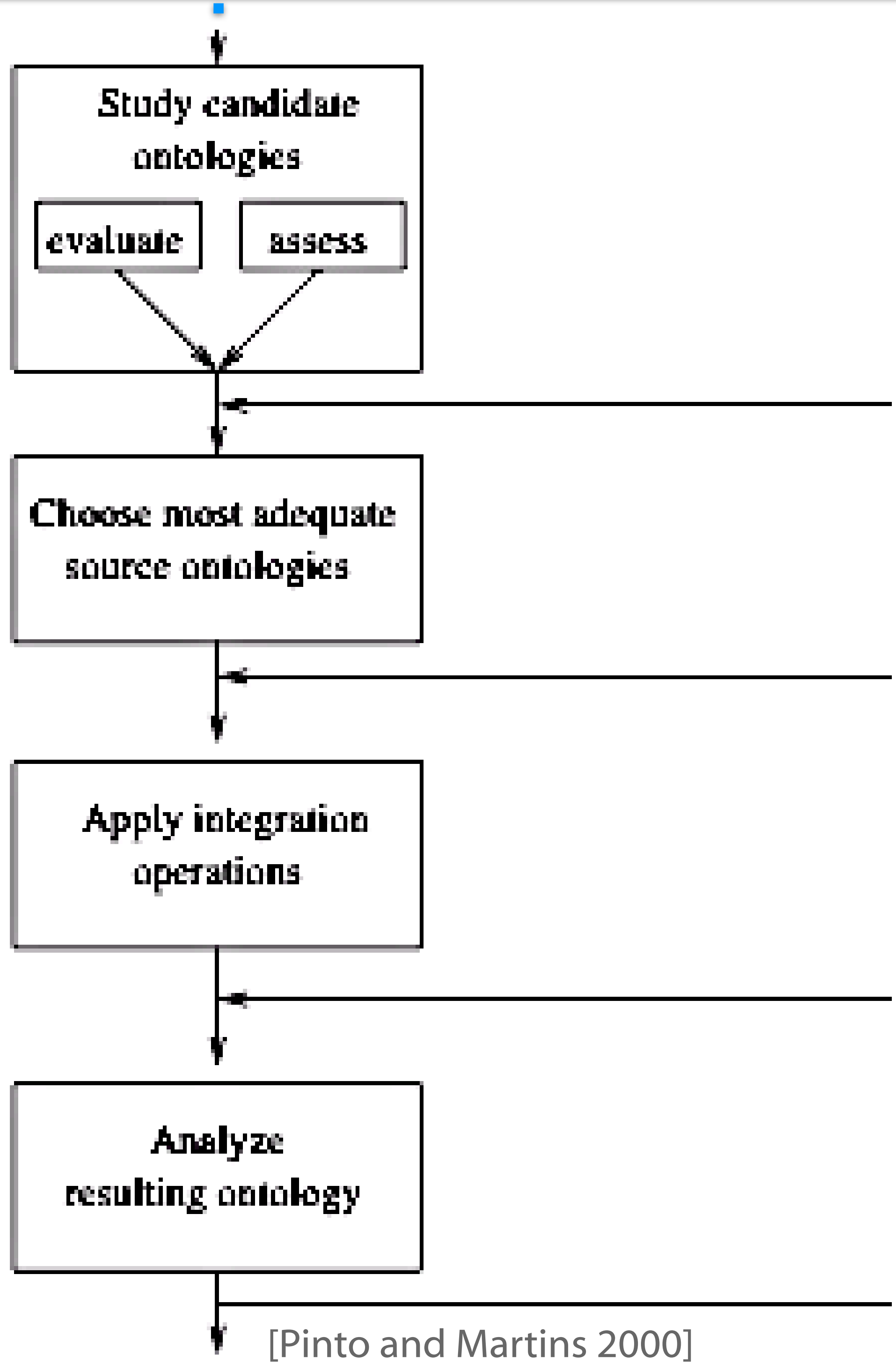
... what knowledge should be represented in each module

[Pinto and Martins 2000]

Integration Process



[Pinto and Martins 2000]



[Pinto and Martins 2000]

Choose Candidate Ontologies

Basic requirements:

- appropriate domain
- is the ontology available?
- formalism paradigms in which the ontology is available
- main assumptions and ontological commitments
- main concepts represented

Additional criteria:

- where is the ontology available?
- at what level is the ontology available?
- what kind of documentation is available
- where is the documentation available?

[Pinto and Martins 2000]

Domain experts evaluate the ontologies in terms of:

- what knowledge is missing
(concepts, classification criteria, relations, etc),
- what knowledge should be removed,
- which knowledge should be relocated,
- which knowledge sources changes should be performed,
- which documentation changes should be performed,
- which terminology changes should be performed,
- which definition changes should be made,
- which practices changes should be made

[Pinto and Martins 2000]

Ontology engineers assess the ontologies in terms of:

- the overall structure of the ontology
- appropriateness classification criteria
- the relation used to structure knowledge
- the naming convention rules
- the quality of the definitions, such as using unified patterns, simple, clear, concise, consistent, complete, correct —semanctically and syntactically—, precise and accurate
- the quality of the documentation
- appropriateness and completeness of the knowledge pieces (entities) represented or included

Taxonomy of features in the **first** stage:

General
- Generality
- Formality
- Development status

Development

Content

Taxonomy of features in the **first** stage:

General

Development

- Knowledge acquisition
  - Quality of knowledge sources
  - Adequacy of knowledge acquisition practices
- Maintenance
  - Is it maintained?
  - Who does maintenance?
  - How is maintenance done?
- Documentation
  - Quality of the documentation available
  - Is the available documentation complete?
- Implementation
  - Language issues

Content

[Pinto and Martins 2000]

Taxonomy of features in the **first** stage:

General

Development

Content

- Level of detail
- Modularity
- Adequacy from the domain expert point of view
- Adequacy from the ontologist point of view

[Pinto and Martins 2000]

Choose Source Ontologies 4

Taxonomy of features in the **second** stage:

Content

- Completeness
- Compatibility
  Terminology of common concepts
  Definitions of common concepts

Criteria to guide integration of knowledge:

- Modularize
- Specialize
- Diversify each hierarchy
- Minimize the semantic distance between sibling concepts
- Maximize relationships between taxonomies
- Standardize names of relations

[Pinto and Martins 2000, Arpirez-Vega et al. 1998]

- Clarity, coherence, extendibility, minimal encoding bias and minimal ontological commitment
- Correctness –semantically and syntactically–, completeness, conciseness, consistency, coherency, expandability, sensitiveness and robustness
- Regular level of detail: no "islands" of exaggerated level of detail

[Pinto and Martins 2000, Gruber 1995, Gómez-Pérez et al. 1995]

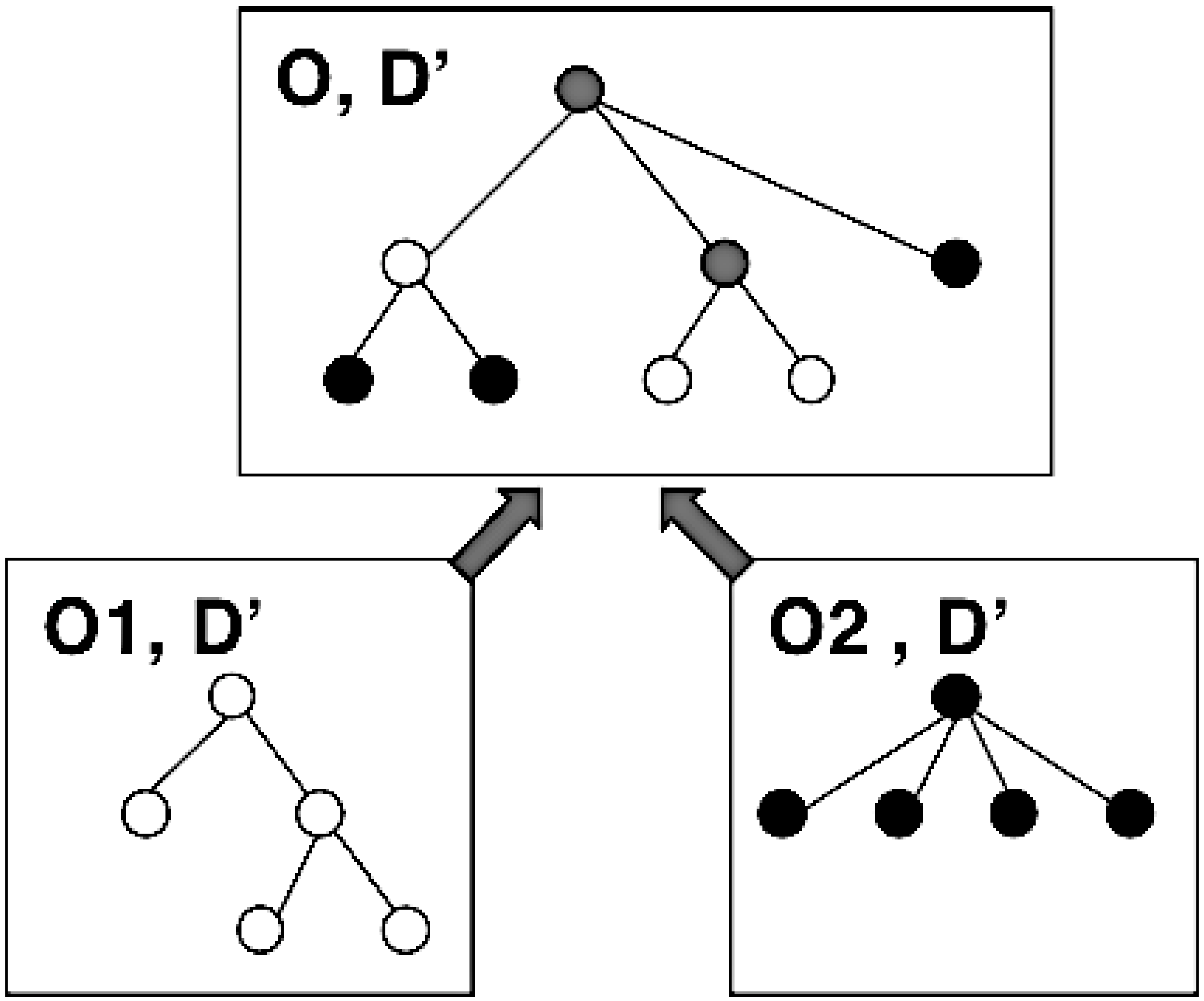
- Ontology Reuse
- Integration
- Merging
- Tools

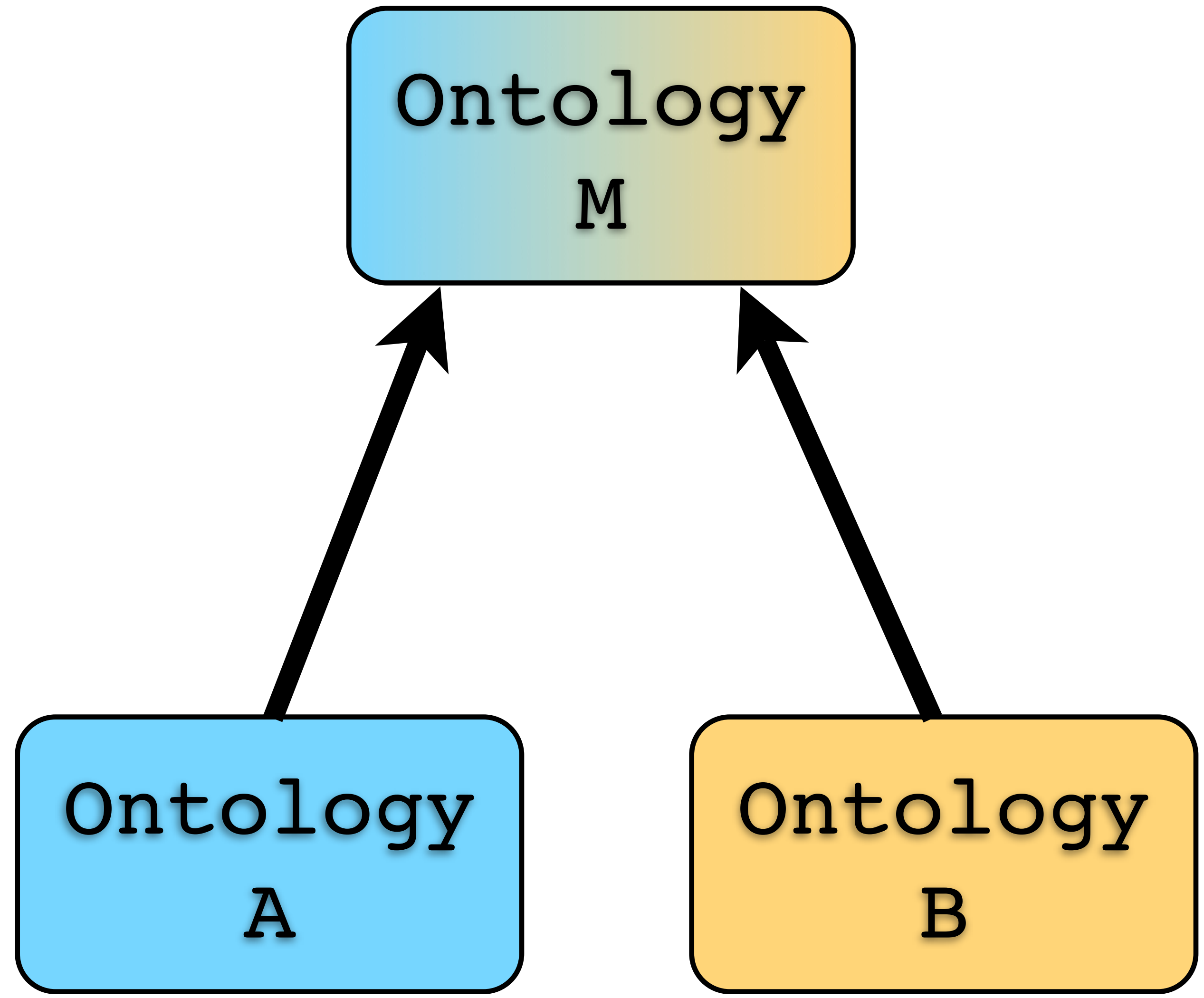
... the creation of a single coherent, consistent and non-redundant ontology made up of two or more source ontologies.

Steps in ontology merging:

- Find the places in the ontologies where they overlap - discover mapping candidates
- Relate concepts that are semantically close via equivalence and subsumption relations
- Check the consistency, coherency and non-redundancy of the result

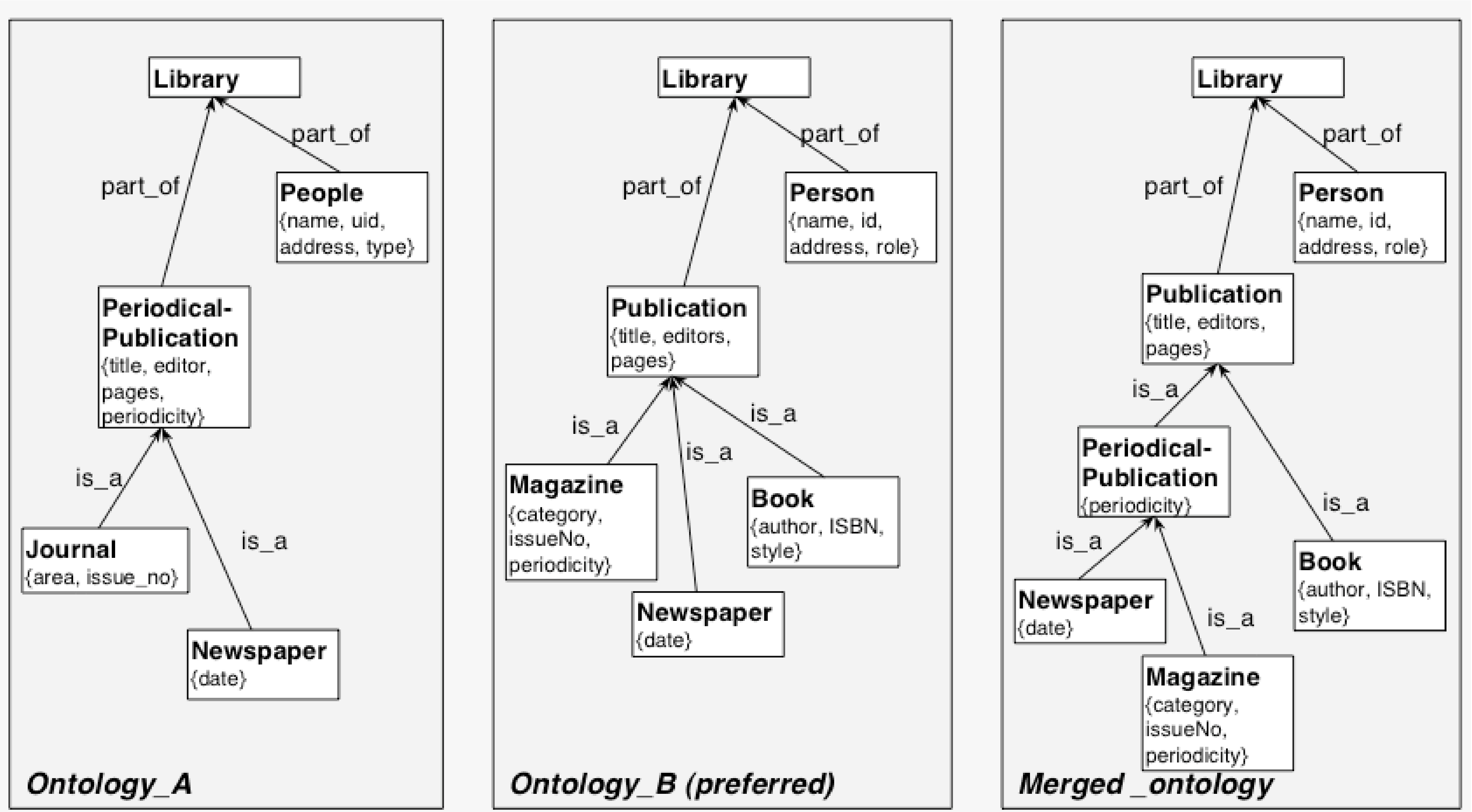
[McGuinness et al. 2000]





[Jakoniene 2003]

Merging Example



[Jakoniene 2003]

Two major architectures for mapping discovery between ontologies exist:

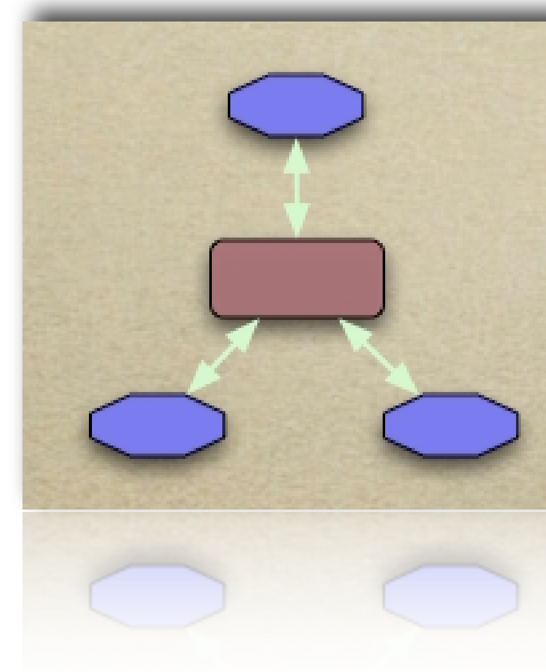
- Using information sources
  - A common reference ontology
  - Lexical information
  - Ontology structure
  - User input
  - External resources
  - Prior matches

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

[Noy 2004, Noy 2005]

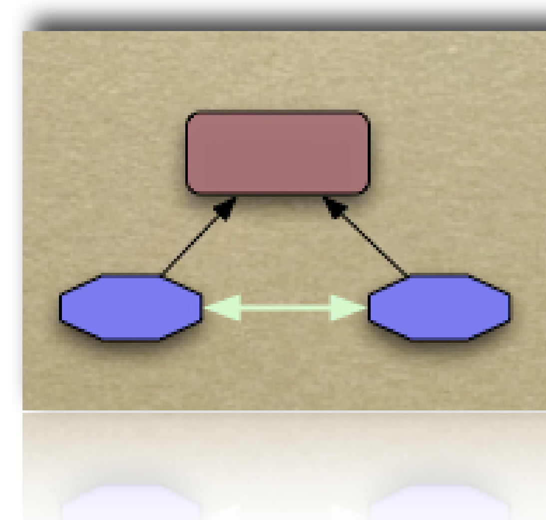
Common Reference Ontology

Upper level or reference ontologies designed to support information integration



- designed in principled way
- provide common reference terminology
- Cyc, SUO, DOLCE

Domain-specific interlingua



- Process Specification Language (PSL)

[Noy 2004, Noy 2005]

String normalization

- upper and lower case
- blanks and delimiters
- diacritics
- stop-words

String distance

- Hamming distance
- Levenshtein distance (edit distance)

Soundex

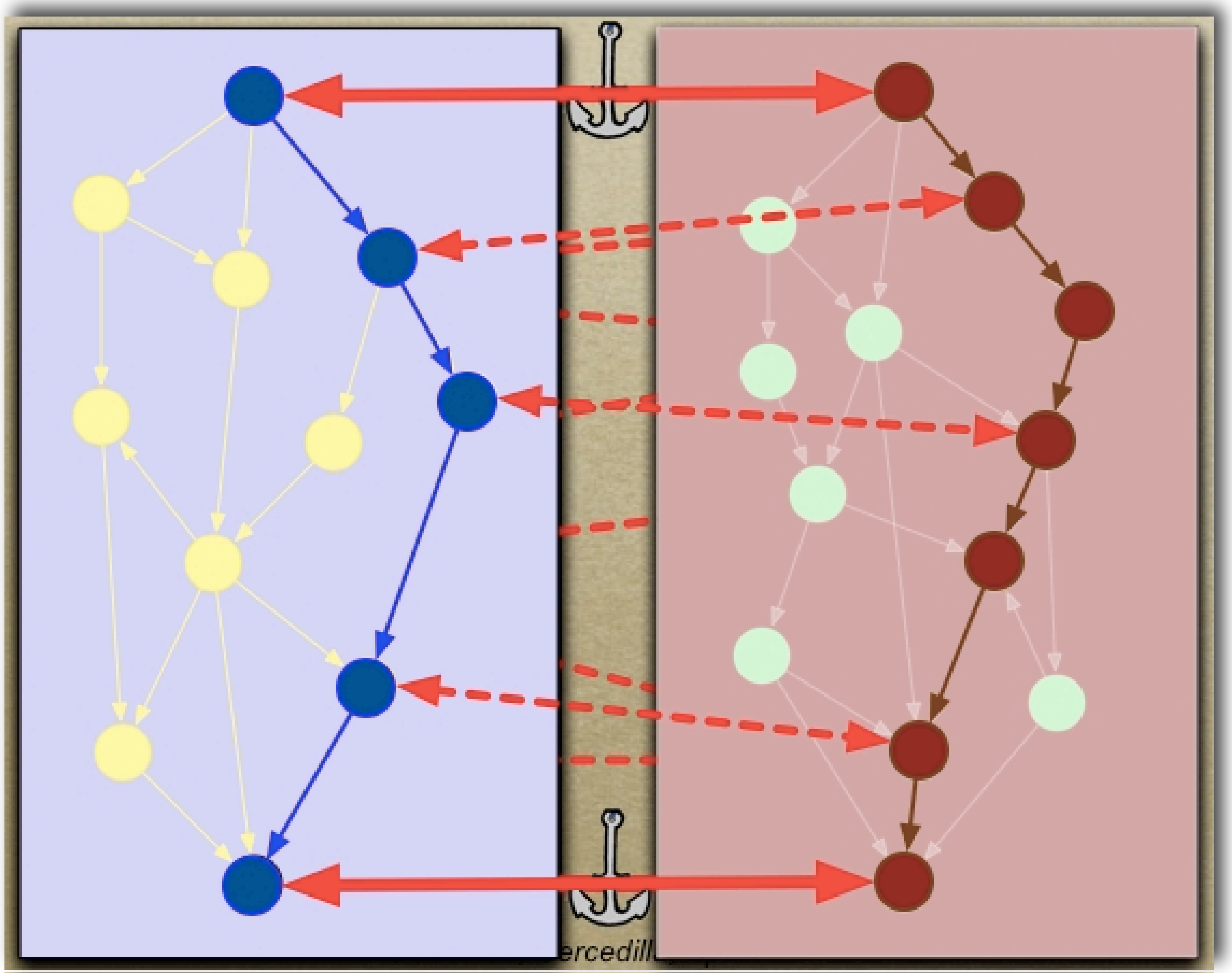
Thesaurus

Natural-language analysis
of concept names and definitions

- splitting composite names
- finding common substrings
- finding the ratio of common words in definitions

Hierarchy information of taxonomies

Treat ontologies as graphs and compare the corresponding subgraphs, e.g. Anchor-Prompt



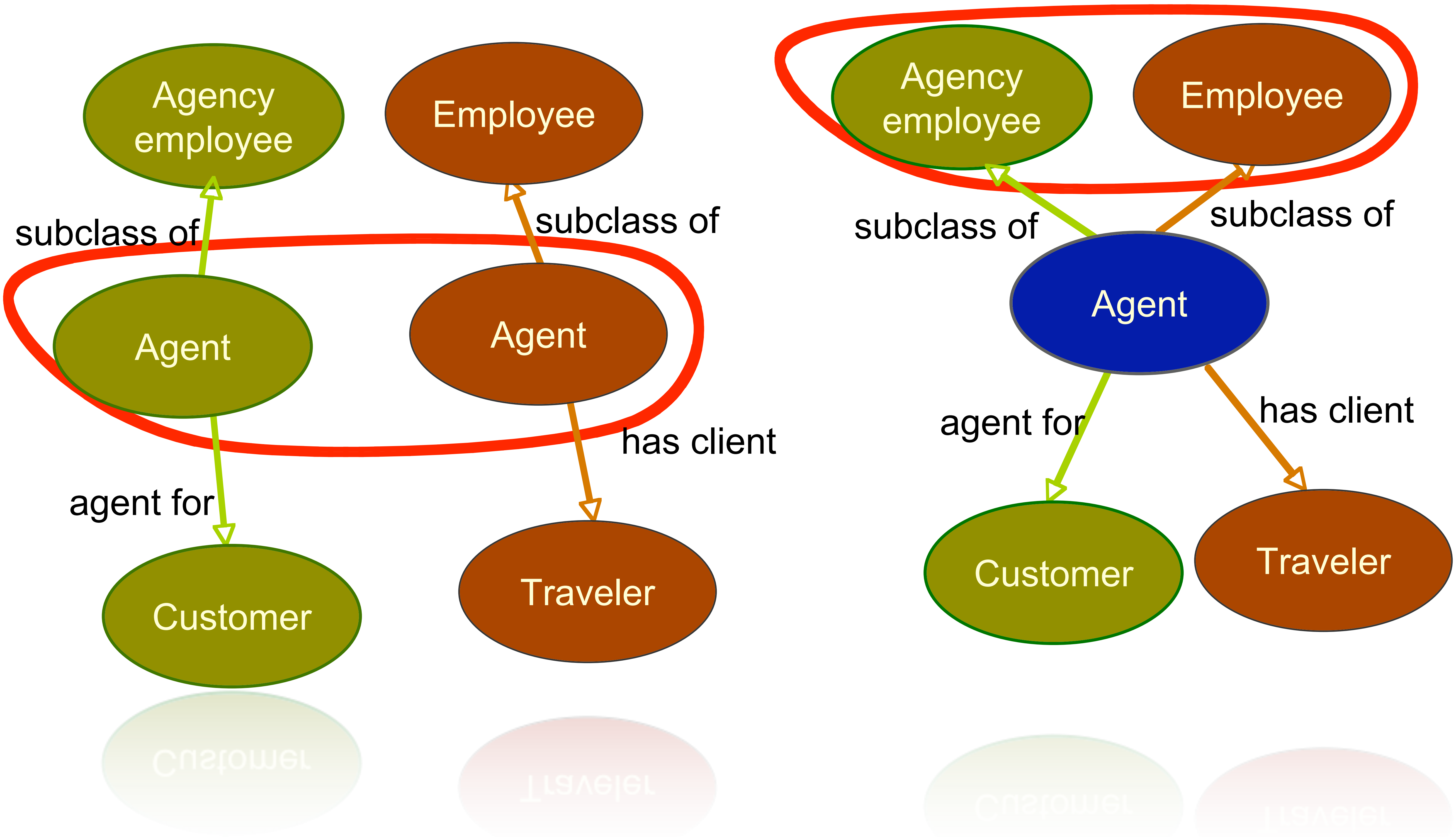
[Noy 2005, Noy and Musen 2001]

Examples for representation of identified mappings:

- As instances in an ontology of mappings
- Defining bridging axioms
  to represent transformations
- Using views to describe mappings
  from a global ontology to local ontologies

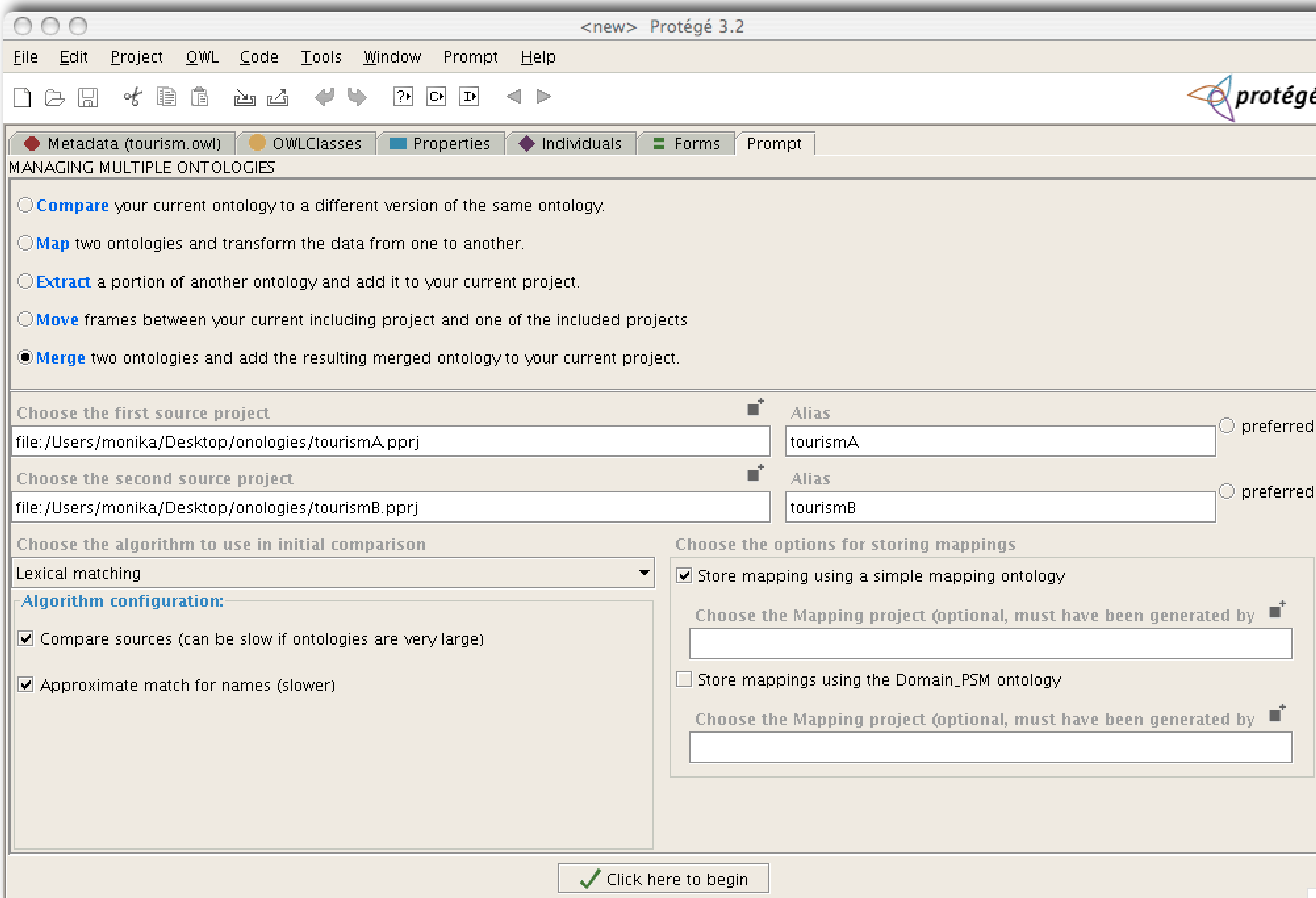
[Noy 2005]

Example: Merge Classes with Prompt



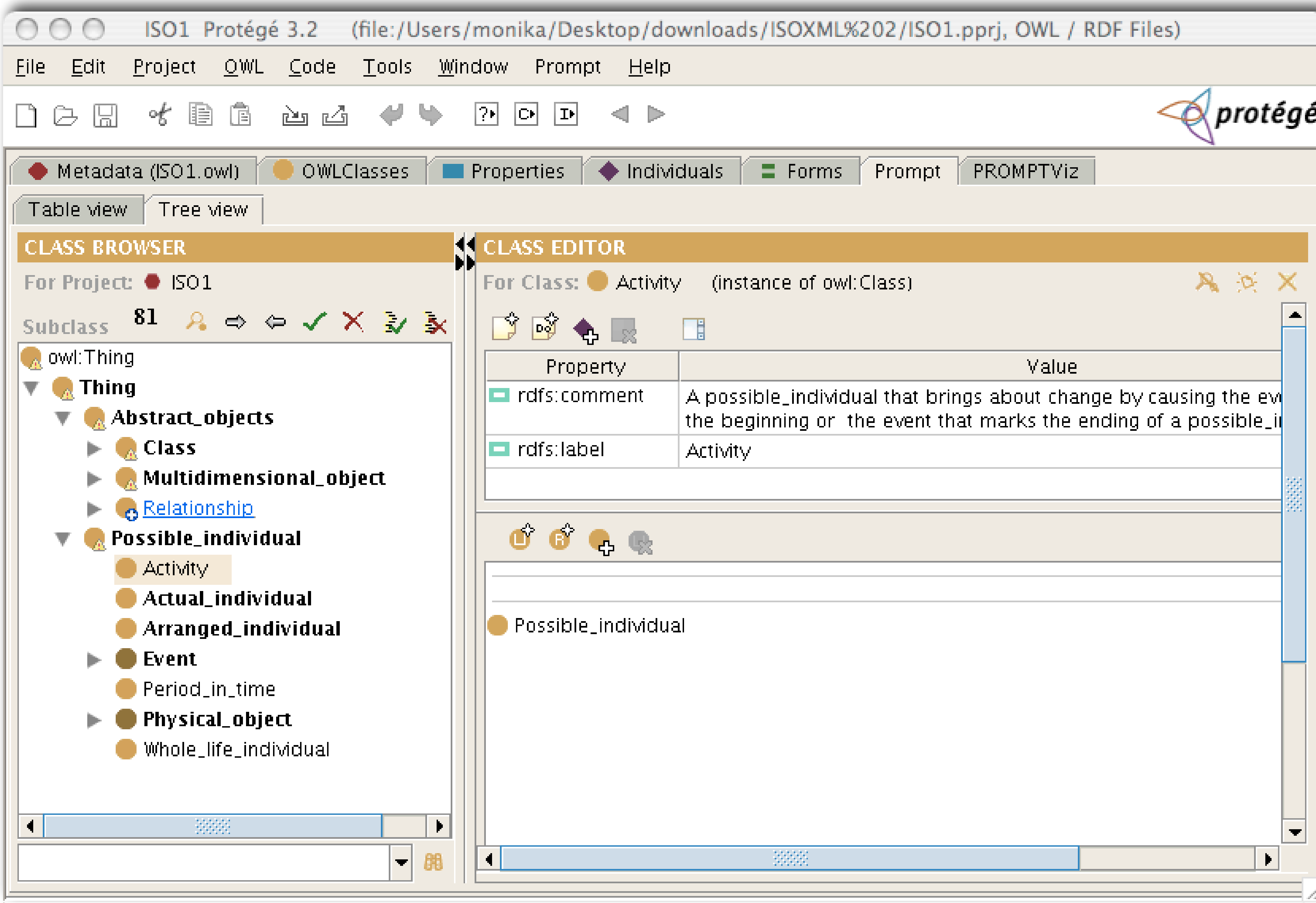
[Noy 2001]

Tools: Protégé and Prompt - Merge



[Protégé 3.2]

Tools: Protégé and Prompt - Compare



[Protégé 3.2]

- Ontology Reuse
- Integration
- Merging
- Tools

Ontolingua
http://www.ksl.stanford.edu/software/ontolingua/

Protégé (current version 3.3)
http://protege.stanford.edu/

OntoStudio
http://www.ontoprise.de

Chimaera
http://ksl.stanford.edu/software/chimaera/

PROMPT
http://protege.stanford.edu/plugins/prompt/prompt.html

WebODE
http://www-sop.inria.fr/acacia/ekaw2000/ode.html

CORE: A Tool for Collaborative Ontology Reuse and Evaluation
http://km.aifb.uni-karlsruhe.de/ws/eon2006/eon2006fernandezetal.pdf

[Arpirez-Vega et al. 1998] Julio C. Arpirez-Vega, Asunción Gómez-Pérez, Adolfo Lozano-Tello, and Helana Sofia Pinto, (ONTO) Agent: An Ontology-Based WWW Broker to Select Ontologies, In Proceedings of ECAI98's Workshop on Application of Ontologies and Problem Solving Methods, pages 16–24, 1998.

[Ehrig, et al. 2004] Marc Ehrig and York Sure, Ontology Mapping - An Integrated Approach, In: Proceedings of the First European Semantic Web Symposium. C. Bussler, J. Davis, D. Fensel, R. Studer, eds, 2004.

[Fernández et al. 1999] Mariano Fernández, M. Asunción Gómez-Pérez, Alejandro P. Sierra, and Juan P. Sierra, Building a Chemical Ontology Using METHONTOLOGY and the Ontology Design Environment. IEEE Intelligent Systems 14(1):37–46, 1999.

[Gómez-Pérez et al. 1995] Asunción Gómez-Pérez, Natalia Juristo, and Juan Pazos, Evaluation and Assessement of the Knowledge Sharing Technology. In Mars, N., ed., Towards Very Large Knowledge Bases, 289–296. IOS Press, 1995.

[Gruber 1995] Thomas R. Gruber, Towards Principles for the Design of Ontologies for Knowledge Sharing. International Journal of Human Computer Studies 43(5/6):907–928, 1995.

[Jakoniene 2003] Vaida Jakoniene. Ontology integration, Lecture Slides, http://www.ida.liu.se/labs/iislab/courses/LW/slides/ontologyIntegration.pdf (checked online 14. Dec. 2006).

[Kalfoglou and Schorlemmer 2003] Yannis Kalfoglou and Marco Schorlemmer. Ontology mapping: the state of the art. The Knowledge Engineering Review, 18(1):1–31, 2003.

[Klein 2001] Michel Klein. Combining and relating ontologies: an analysis of problems and solutions. In Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01), Workshop: Ontologies and Information Sharing, Seattle, USA, 2001.

[McGuinness et al. 2000] Deborah L. McGuinness, Richard Fikes, James Rice, and Steve Wilder. An environment for merging and testing large ontologies. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, KR2000: Principles of Knowledge Representation and Reasoning, pp. 483–493, San Francisco, Morgan Kaufmann, 2000.

[Noy 2005] Natalya. F. Noy. Ontology Mapping and Alignment, Tutorial, The Third Summer School on Ontological Engineering and the Semantic Web (SSSW'05), 2005.

References & Resources

[Noy 2004] Natalya. F. Noy, Semantic Integration: A Survey Of Ontology-Based Approaches. SIGMOD Record, Special Issue on Semantic Integration,33(4), 2004.

[Noy 2001] Natalya F. Noy, Managing Multiple Ontologies in Protégé-2000, In Proceedings of the Fifth International Protégé-2000 Workshop, Newcastle, England, July 2001.

[Noy and Musen 2001] Natalya. F. Noy, Mark A. Musen, 2001. Anchor-PROMPT: Using non-local context for semantic matching. In Workshop on Ontologies and Information Sharing at the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001). Seattle, WA, 2001.

[Pinto, et al. 1999] Helena S. Pinto, Asunción Gómez-Pérez, A.; and Joao P. Martins. Some Issues on Ontology Integration. In: IJCAI99's workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends, 1999.

[Pinto and Martins, 2000] Helena S. Pinto and Joao P. Martins, Reusing Ontologies, In Proceedings of AAAI2000 Spring Symposium Series, Workshop Bringing Knowledge to Business Processes: AAAI Technical Report SS-00-03, Menlo Park, California, USA, AAAI Press, pages 77-84, 2000.

[Protégé 3.2] Stanford Medical Informatics, Stanford University, CA, Protégé, http://protege.stanford.edu/download/old-releases/3.2 (checked online 6. Jan. 2008).

[Su 2004] Xiaomeng Su, Semantic Enrichment for Ontology Mapping, dr.ing. thesis, NTNU 2004:116, ISBN 82-471-6353-1, ISSN 1503-8181, 219 p. Disputas: 15 Dec. 2004. Advisor: Arne Sølvberg (IS). 2004.