Semi-Automatic Information and Knowledge Systems

:
Merging & Integration and Exercises

Monika Lanzenberger

- Exercises
- Ontology Reuse
- Integration
- Merging
- Tools

- 15 topics are available:
  different ontology mapping and merging tools (some with graphical interfaces) and related work

- Select one topic
  either alone or together with another colleague
- Investigate and read the given references
- Search for additional material
- Test ontologies are available on the web:
  http://www.aifb.uni-karlsruhe.de/WBS/meh/foam/ontologies.htm

---

- Prepare a 15 minutes talk
  (Deutsch or English)
- Teams prepare a 30 minutes talk

- Send the presentation material
  10 days before your talk to ML
- Include the feedback and produce
  a final version of the presentation material

- Participate in the discussions

Framework for Ontology Alignment and Mapping

- University of Karlsruhe, Germany
- Fully or semi-automatically aligns two or more OWL ontologies
- Heuristics (similarity) of the individual entities (concepts, relations, and instances)
- Java application, web service and Protégé plug-in available

Resources:

- http://www.aifb.uni-karlsruhe.de/WBS/meh/foam/
- http://www.aifb.uni-karlsruhe.de/WBS/meh/publications/ehrig04similarityPAKM.pdf
- http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-156/paper11.pdf
- http://www.moemais.ufam.edu.br/plugin.php
- http://protege.stanford.edu/download/download.html
- http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-156/paper11.pdf

Deliverables:

- Run FOAM with the tourism ontologies (application, web service, and plug-in)
- Search for additional material / literature
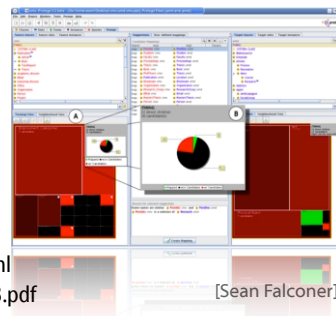- Prepare slides and a demo for your short presentation

---

Manage multiple ontologies

- Stanford University, CA, USA
- Plug-in for Protégé
- Manages ontologies: map, compare, merge, move, extract
- Creates visual representations of the differences between two versions of an ontology

[thechiselgroup.org]

Resources:

- http://protege.stanford.edu/plugins/prompt/prompt.html
- http://smi.stanford.edu/smi-web/reports/SMI-2003-0973.pdfChisel logo
- http://www.thechiselgroup.org/promptviz
- http://protege.stanford.edu/download/download.html

Deliverables:

- Install Protégé and test PROMPT and PROMPTViz
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is the PROMPT suite? What is PromptViz?
  How do they work in principle and in practice?
  Describe strengths and shortcomings of both tools.

Manage multiple ontologies

- Stanford University, CA, USA
- Plug-in for Protégé
- Map, compare, merge, move, extract
- Visual representation

[Sean Falconer]

Resources:

- http://protege.stanford.edu/plugins/prompt/prompt.html
- http://smi.stanford.edu/smi-web/reports/SMI-2003-0973.pdf
- http://protege.cim3.net/cgi-bin/wiki.pl?PluginsForPrompt
- http://www.cs.uvic.ca/~seanf/cogz/cogz.html
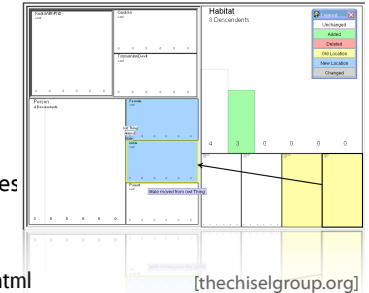- http://protege.stanford.edu/download/download.html

Deliverables:

- Install Protégé and run the CogZ Plug-in
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is Prompt? What is CogZ? How to develop plug-ins for Prompt?
  How does CogZ work in principle and in practice? What is the benefit of the visualization?
  Describe strengths and shortcomings of the tool.

Risk Minimization based Ontology Mapping

- Tsinghua University, Beijing, China
- Maps entities (n:1, 1:null, null:1, and 1:1)
- Each entity e_{i1} (O_1) can be classified to one 'class' e_{i2} (O_2).
- Classification as a special case of Bayesian decision problem
- loss function
- Find minimal risk for each sample

Resources:

- http://keg.cs.tsinghua.edu.cn/project/RiMOM/
- http://www.websemanticsjournal.org/ps/pub/showDoc.Fulltext/document.pdf?lang=en&doc=2006-2&format=pdf&compression=
- http://www.dit.unitn.it/~p2p/OM-2006/17-RiMOM-OAEI'06.pdf

Deliverables:

- Install and run RiMOM
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is RiMOM? How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool. Outline an application scenario.

⑤ OLA

OWL Lite Alignment, VisOn

- University of Montreal, Canada
- Java tool for ontology alignment
- Graph-based approach to represent ontologies

OLA
[OWL Lite Alignment]

Resources:

- http://www.iro.umontreal.ca/%7Eowlola/userguide.html
- http://www.iro.umontreal.ca/%7Eowlola/pdf/align-compet-EON.pdf
- http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-156/paper15.pdf

Deliverables:

- Install the relevant tools
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is VisOn? How is the alignment done? What is the benefit of the visualization?
  How does OLA works in principle and in practice?
  Describe strengths and shortcomings of the tool.

⑥ Chimaera

Supports users in creating and maintaining distributed ontologies on the web

- Stanford University, CA, USA
- Merging multiple ontologies
- Diagnosing individual or multiple ontologies

Chimæra

Resources:

- http://ksl.stanford.edu/software/chimaera/
- http://www.ksl.stanford.edu/people/dlm/papers/kr2000-camera-ready-copy-withcitation.doc

Deliverables:

- Install and run Chimaera
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is Chimaera?
  How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool.
  Is there ongoing work or is it outdated?

⑦ Falcon-AO

Automatic ontology matching tool

- Southeast University, Nanjing, China
- For OWL Lite/DL ontologies (schema-level)
- Three matchers integrated in Falcon-AO:
  V-Doc, String (includes edit distance and I-Sub) and GMO + PBM for large-scale ontologies (partition-based block matching).

Resources:

- http://www.dit.unitn.it/~p2p/OM-2006/11-Falcon-OAEI'06.pdf
- http://xobjects.seu.edu.cn/project/falcon/matching/resources.html
- http://xobjects.seu.edu.cn/project/falcon/pub/Falcon-AO%20-%20Aligning%20Ontologies%20with%20Falcon.pdf
- http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-156/paper13.pdf

Deliverables:

- Install Falcon-AO and run it
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is Falcon-AO?
  How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool.

⑧ COMA++

Schema and Ontology Matching

- University Leipzig, Germany
- Identifies semantic correspondences between metadata structures or models such as database schemas, XML message formats, and ontologies
- Graphical interface

Resources:

- http://www.dit.unitn.it/~p2p/OM-2006/9-coma-OAEI'06.pdf
- http://dbs.uni-leipzig.de/file/comaplusplus.pdf
- http://dbs.uni-leipzig.de/Research/coma_index.html
- http://dbs.uni-leipzig.de/Research/coma.html

Deliverables:

- Install COMA++ and run it
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is COMA++? What is the benefit of the graphical interface?
  How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool.
  Outline an application scenario.

⑨ AUTOMS

Automating Ontology Mapping through Synthesis of Approaches
- University of the Aegean, Samos, Greece
- Integrates several matching methods: lexical, semantic and structural
- Based on HCONE-merge and COCLU (COmpression-based CLUstering))
- Concepts, instances and properties are investigated for similarity

Resources:
- http://www.dit.unitn.it/~p2p/OM-2006/8-automs-OAEI'06.pdf
- http://www.icsd.aegean.gr/incosys_old/Projects/AUTOMS/OAEI/
- http://www.websemanticsjournal.org/ps/pub/showDoc.Fulltext/document.pdf?lang=en&doc=2005-33&format=pdf&compression=
- http://www.ifs.tuwien.ac.at/~mlanzenberger/teaching/WS0607/SAIKS/stuff/automs.zip

Deliverables:
- Install and run AUTOMS
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is AUTOMS? Describe the different matching methods.
  How does AUTOMS work in principle and in practice?

⑩ OWL-CTXMatch

Framework for Ontology Alignment and Mapping
- Silesian University of Technology, Gliwice, Poland
- Java Tool for OWL DL ontologies
- Uses WordNet as a source of lexical and domain knowledge
- Maps classes and properties

Resources:
- http://www.dit.unitn.it/~p2p/OM-2006/15-OWL-CtxMatch-OAEI'06.pdf
- http://www-zo.iinf.polsl.gliwice.pl/~niedbyk/oaei2006/
- For more information on the installation procedure ask ML

Deliverables:
- Install and run OWL-CTXMatch
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is CTX-Match? What is OWL-CTXMatch?
  How does OWL-CTXMatch work in principle and in practice?
  Describe strengths and shortcomings of the tool.

⑪ FCA-Merge

Formal Concept Analysis
- Merges ontologies bottom-up based on common instances
- Concepts are generated, explored and transformed to the merged ontology

Resources:
- http://www.aifb.uni-karlsruhe.de/WBS/gst/papers/2001/IJCAI01.pdf

Deliverables:
- Search for additional material / literature
- Prepare slides for your short presentation
  What is FCA-Merge?
  How does it work in principle?
  Describe strengths and shortcomings of this approach.
  Outline an application scenario.

⑫ CMS

CROSI Mapping System
- University of Southampton, UK & Hewlett Packard Laboratories
- Uses name matchers (e.g. WordNet) and semantic matchers (structure- and intension-aware)
- Identifies corresponding entities

Resources:
- http://www.aktors.org/crosi/deliverables/downloads/
- http://www.aktors.org/crosi/deliverables/pdfs/kcap05kh.pdf
- http://www.aktors.org/crosi/
- http://eprints.ecs.soton.ac.uk/11717/01/crosi-final-report.pdf

Deliverables:
- Install and run CMS
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is CMS?
  How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool.

Framework for automatically aligning OWL ontologies

- ISTI-Italian National Research Council, Pisa, Italy
- Identifies mapping candidates together with their weights
- Uses different classifiers to estimate the quality of a mapping

Resources:

- http://homepages.cwi.nl/~troncy/
- http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-156/paper14.pdf
- http://faure.isti.cnr.it/~straccia/download/papers/SWAP05/SWAP05.pdf

Deliverables:

- Install and run oMap
- Search for additional material / literature
- Prepare slides and a demo for your short presentation
  What is oMap? Which classifiers are applied?
  How does it work in principle and in practice?
  Describe strengths and shortcomings of the tool.
  Outline an application scenario.

Exploiting the structure of background knowledge used in ontology matching

Zharko Aleksovski, Warner ten Kate, and Frank van Harmelen

Resources:

- http://www.dit.unitn.it/~p2p/OM-2006/2-Aleksovski-TP-OM'06.pdf

Deliverables:

- Search for relevant additional material / literature
- Prepare slides for your short presentation
  Why should we use background knowledge?
  Discuss the results of the paper: Describe strengths and shortcomings.
  Outline an application scenario.

Towards understanding the needs of cognitive support for ontology mapping

Sean M. Falconer, Natalya F. Noy, and Margaret-Anne Storey

Resources:

- http://www.dit.unitn.it/~p2p/OM-2006/3-Falconer-TP-OM'06.pdf

Deliverables:

- Search for relevant additional material / literature
- Prepare slides for your short presentation
  How can we support the ontology mapping process? How to apply visualization methods?
  Discuss the results of the paper: Describe strengths and shortcomings of current user interfaces.

- Exercises
- 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

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)
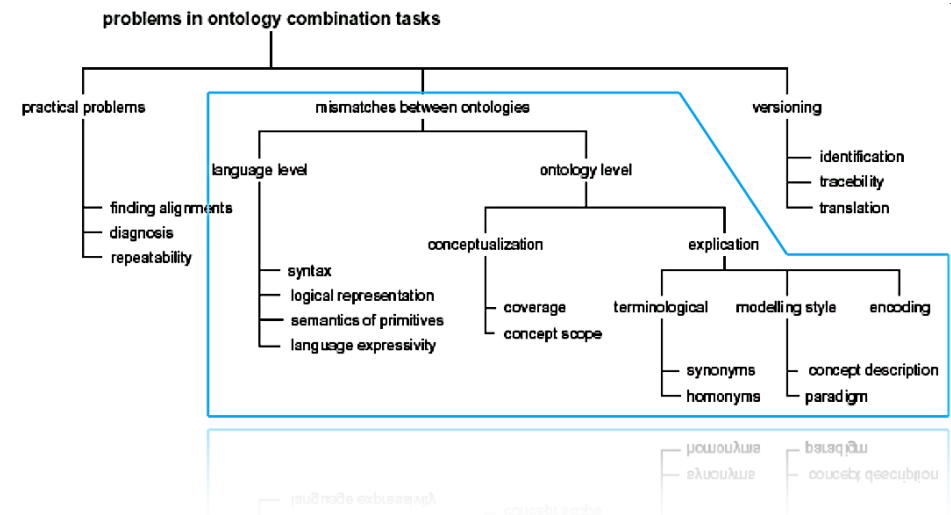
Syntax (e.g., Class definition):

- `<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.)

[Klein 2001, Su 2004]   ML

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

[Klein 2001, Su 2004]   ML

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]   ML

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]   ML

- Exercises
- Ontology Reuse
- Integration
- Merging
- Tools

Integration process takes place along the entire ontology building life cycle
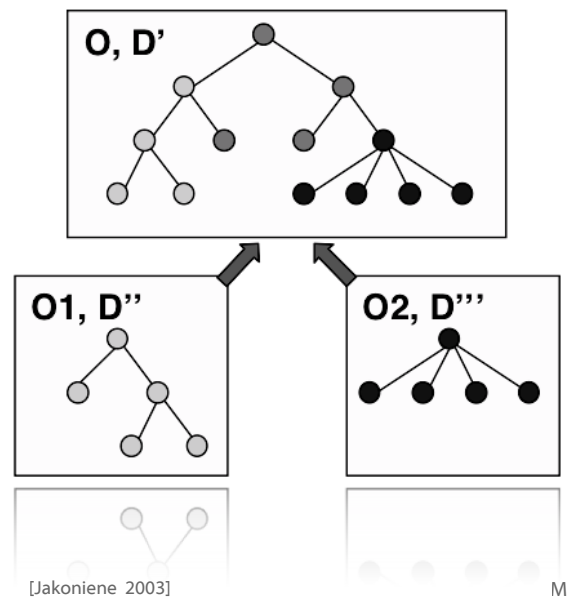
METHONTOLOGY
Stages of the building process:
- Specification
- Conceptualization
- Formalization
- Implementation
- Maintenance

… integration should begin as early as possible.

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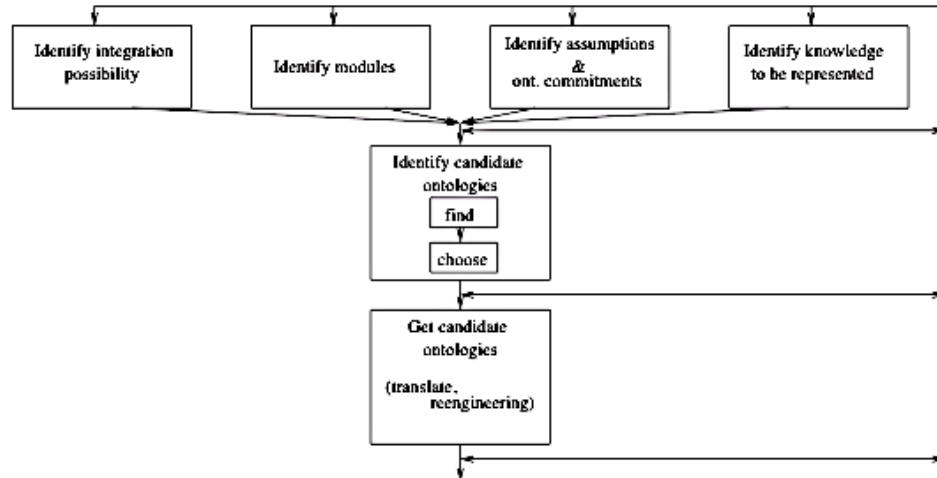$O, O_1,$ and O_2
Different domains:
D', D", and D"

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

Choose Candidate Ontologies   35

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?
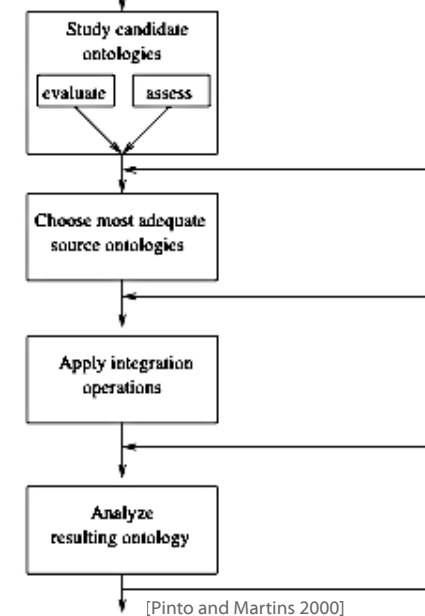
Evaluate Candidate Ontologies   36

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

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

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

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

- 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

- Exercises
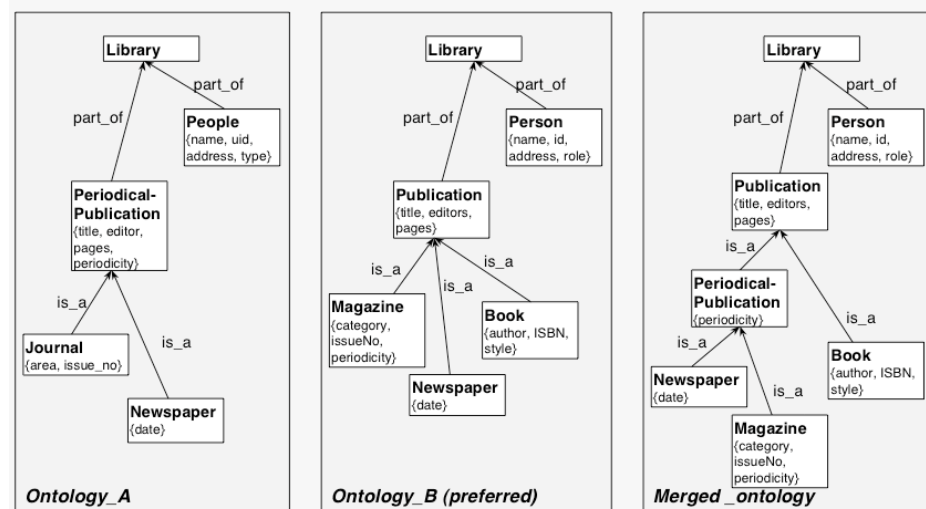- Ontology Reuse
- Integration
- Merging
- Tools

...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]   ML

[Jakoniene 2003]   ML

[Jakoniene 2003]   ML
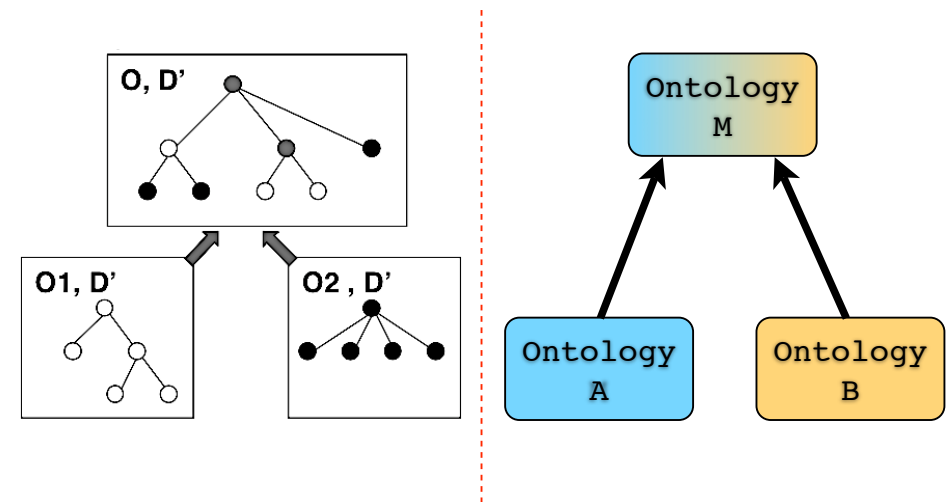
Two major architectures for mapping discovery between
ontologies exist:

- 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

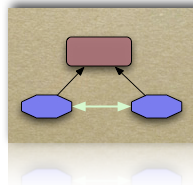[Noy 2004, Noy 2005]   ML

Upper level or reference ontologies designed to support information integration

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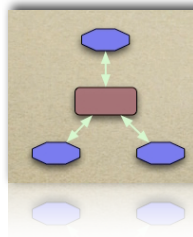
Domain-specific interlingua

- Process Specification Language (PSL)

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
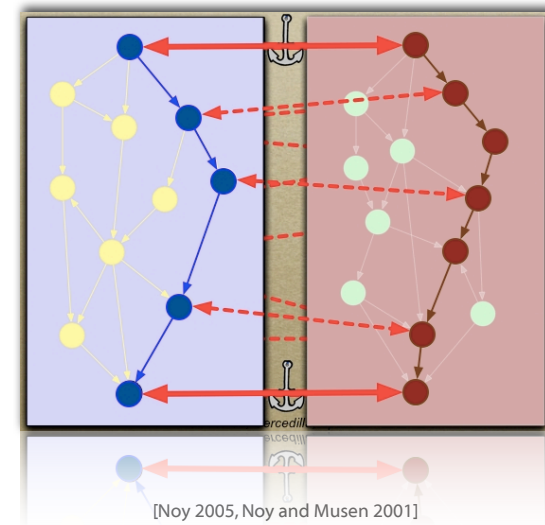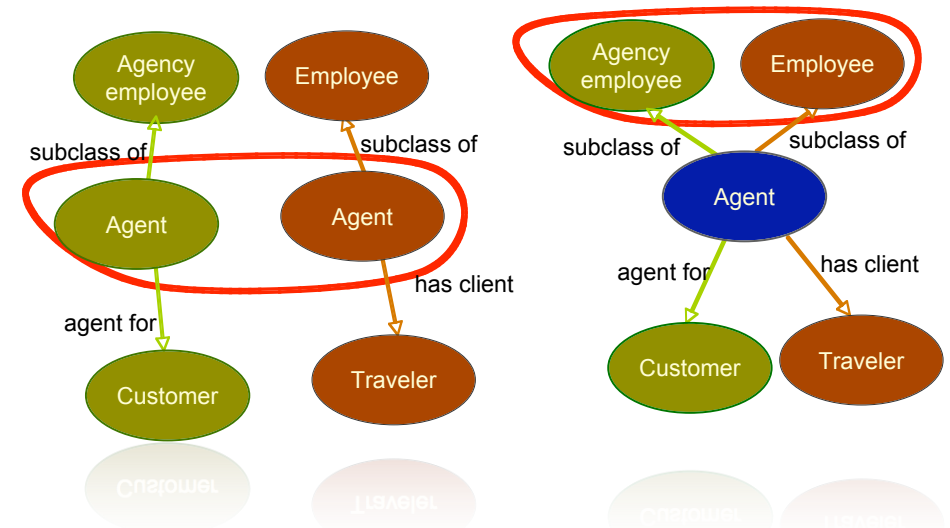
Examples for representation of identified mappings:

- As instances in an ontology of mappings

- Defining bridging axioms
  to represent transformations

- Using views to describe mappings
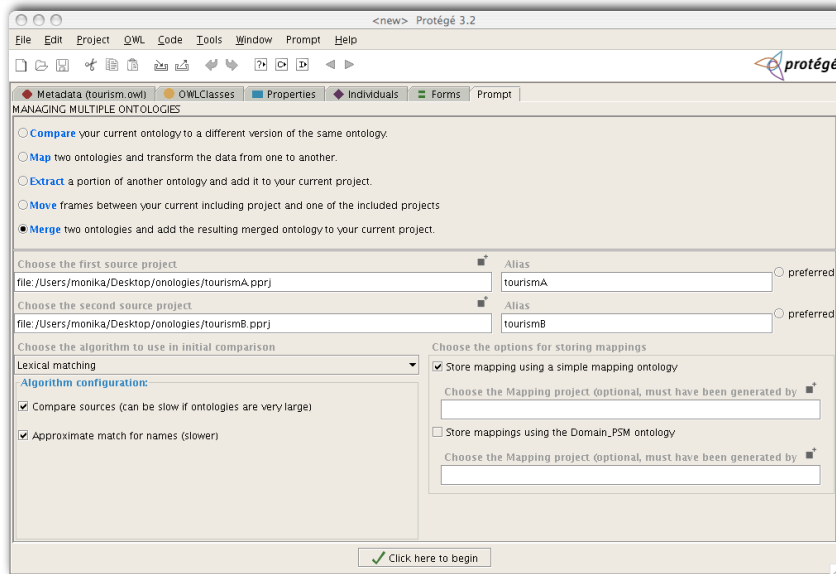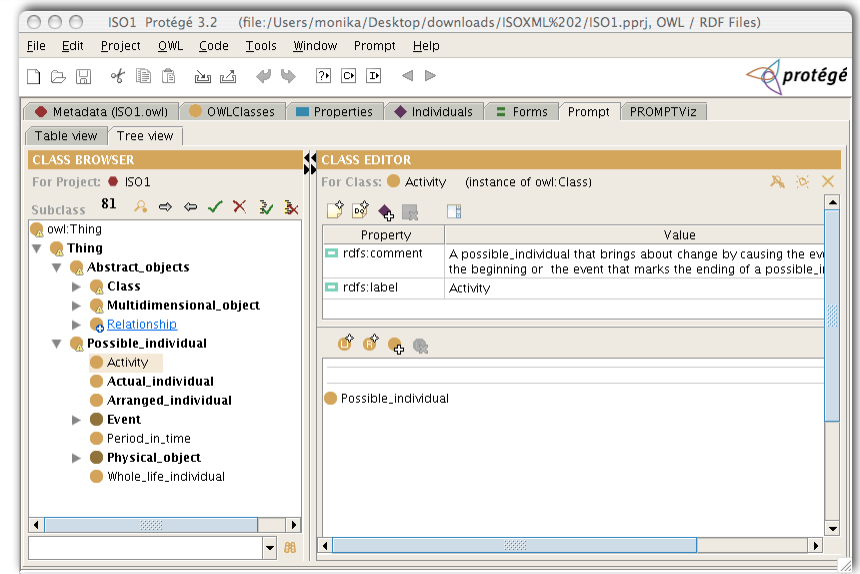 from a global ontology to local ontologies

- Exercises
- Ontology Reuse
- Integration
- Merging
- Tools

ML

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

Protégé (current version 3.2)
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

ML

[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.

ML

[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.

ML

[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/download.html (checked online 14. Dec. 2006).

[Sean Falconer] Sean Falconer, The CogZ Project, University of Victoria, Computer Human Interaction & Software Engineering Lab, http://www.cs.uvic.ca/~seanf/cogz/cogz.html (checked online 14. Dec. 2006).

[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.

[thechiselgroup.org] University of Victoria, Computer-Human Interaction and Software Engineering lab, PromptViz, http://www.thechiselgroup.org/promptviz (checked online 14. Dec. 2006).