Ontology Learning

Ícaro Medeiros

CIn - UFPE

September 30, 2008

Outline

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

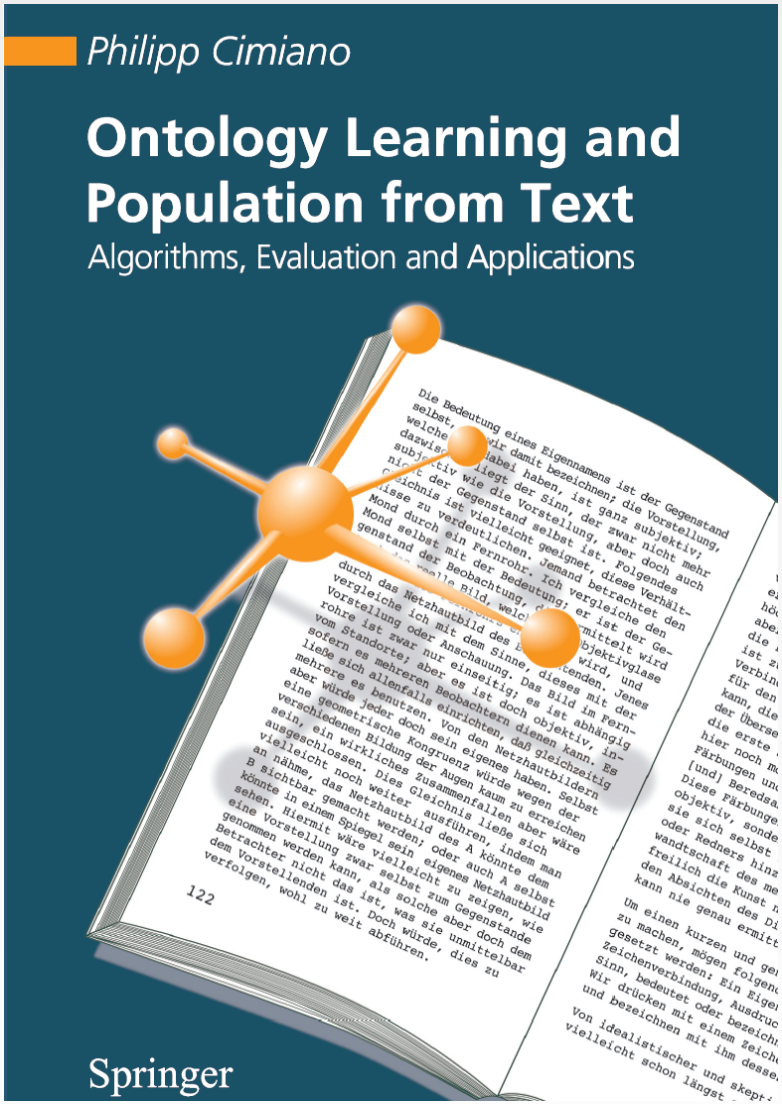
Sections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Too many names, the same subject

- Ontology
  - Extraction
  - Emergence
  - Generation
  - Acquisition
  - Discovery
  - Population
  - Enrichment

Ontology Learning!



(Cimiano, 2006)

WHAT is Ontology Learning (OL)?

- Methods and techniques for (OntoSum, 2008):
  - Building an ontology from scratch
  - Enriching, or adapting an existing ontology
- Extract concepts and relations to form an ontology (Wikipedia, 2008a)
- OL is a semi-automatic task of information extraction

What is Ontology Learning for? (WHY)

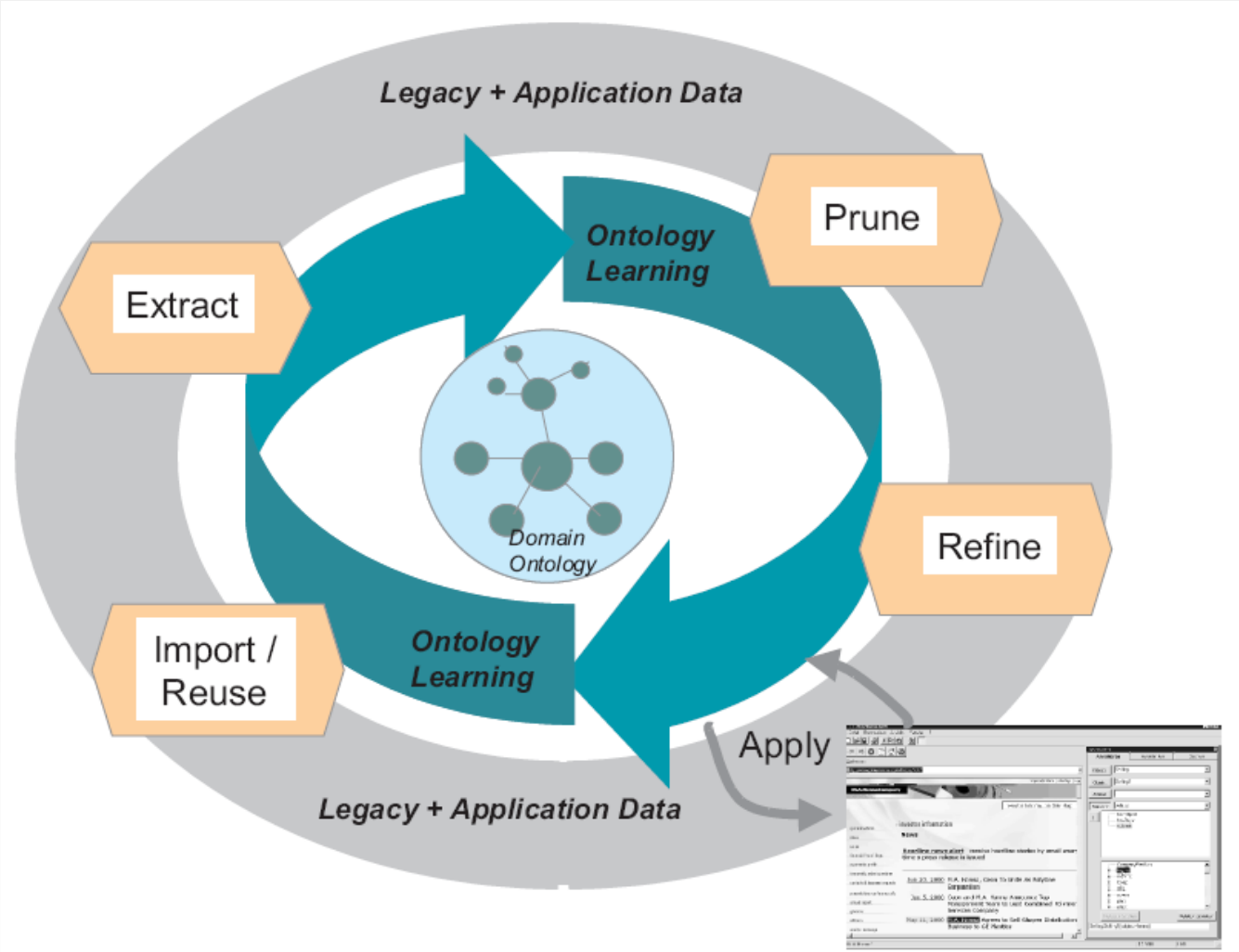
- Problems in Ontology Engineering (OE) (Maedche and Staab, 2001):
  - Can you develop an ontology fast? (time)
  - Is it difficult to build an ontology? (difficulty)
  - How do you know that you've got the ontology right? (confidence)
- OL can overcome these problems, specially the Knowledge Acquisition bottleneck

Information Sources

- Relevant text (Web documents mainly)
- Web document schemata (XML, DTD, RDF)
- Databases on the Web
- Dictionaries
- Semi-structured documents
- Personal Wikis, e-mail/file folders
- Existing Web ontologies

OE Cycle (Maedche and Staab, 2001)

- OL is not only the task of extraction



Sections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

How to Learn Ontologies?

- Natural Language Processing
- Dictionary Parsing
- Statistical Analysis
- Machine Learning
- Hierarchical Concept Clustering
- Formal Concept Analysis (Lattices)

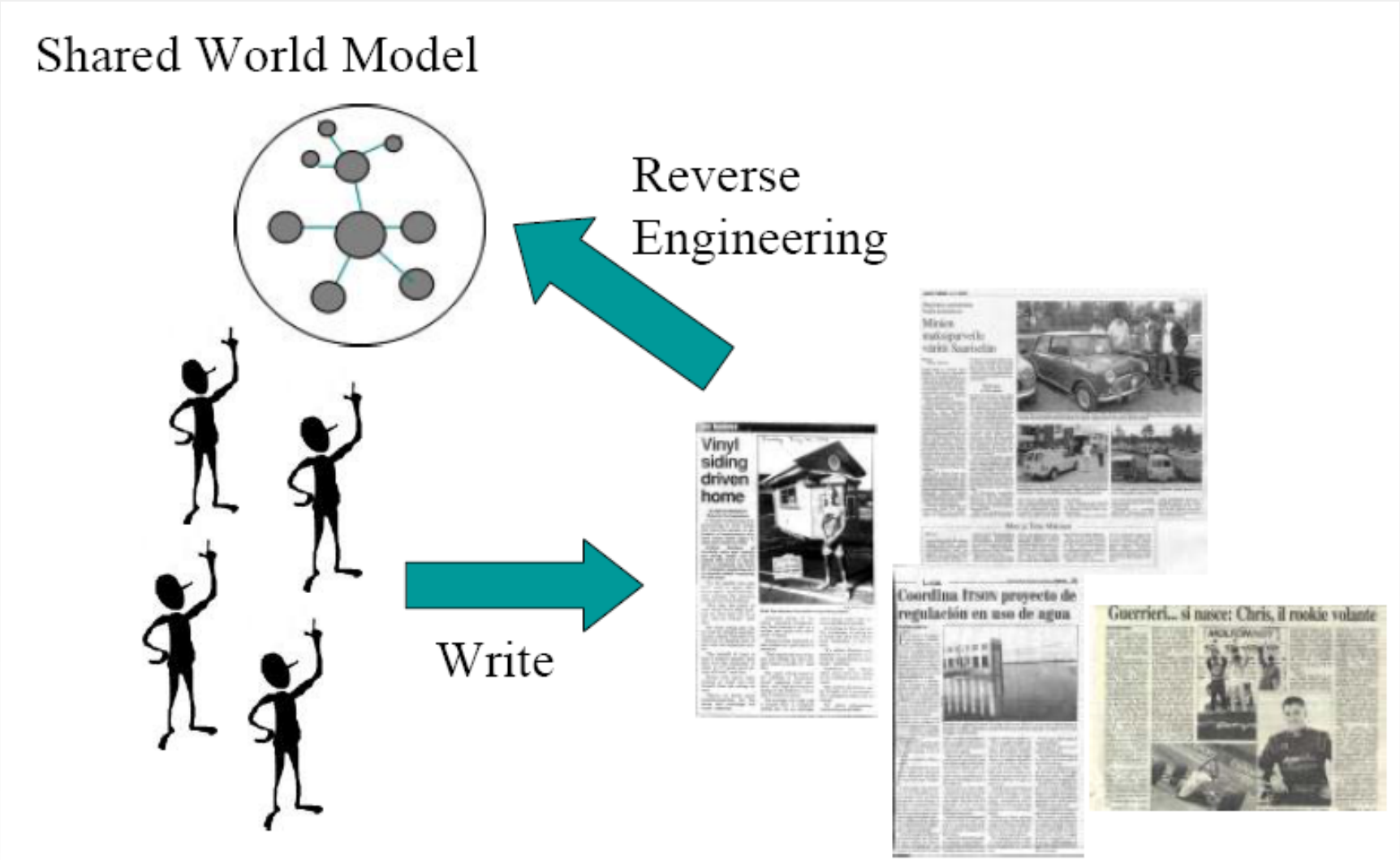
Subsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

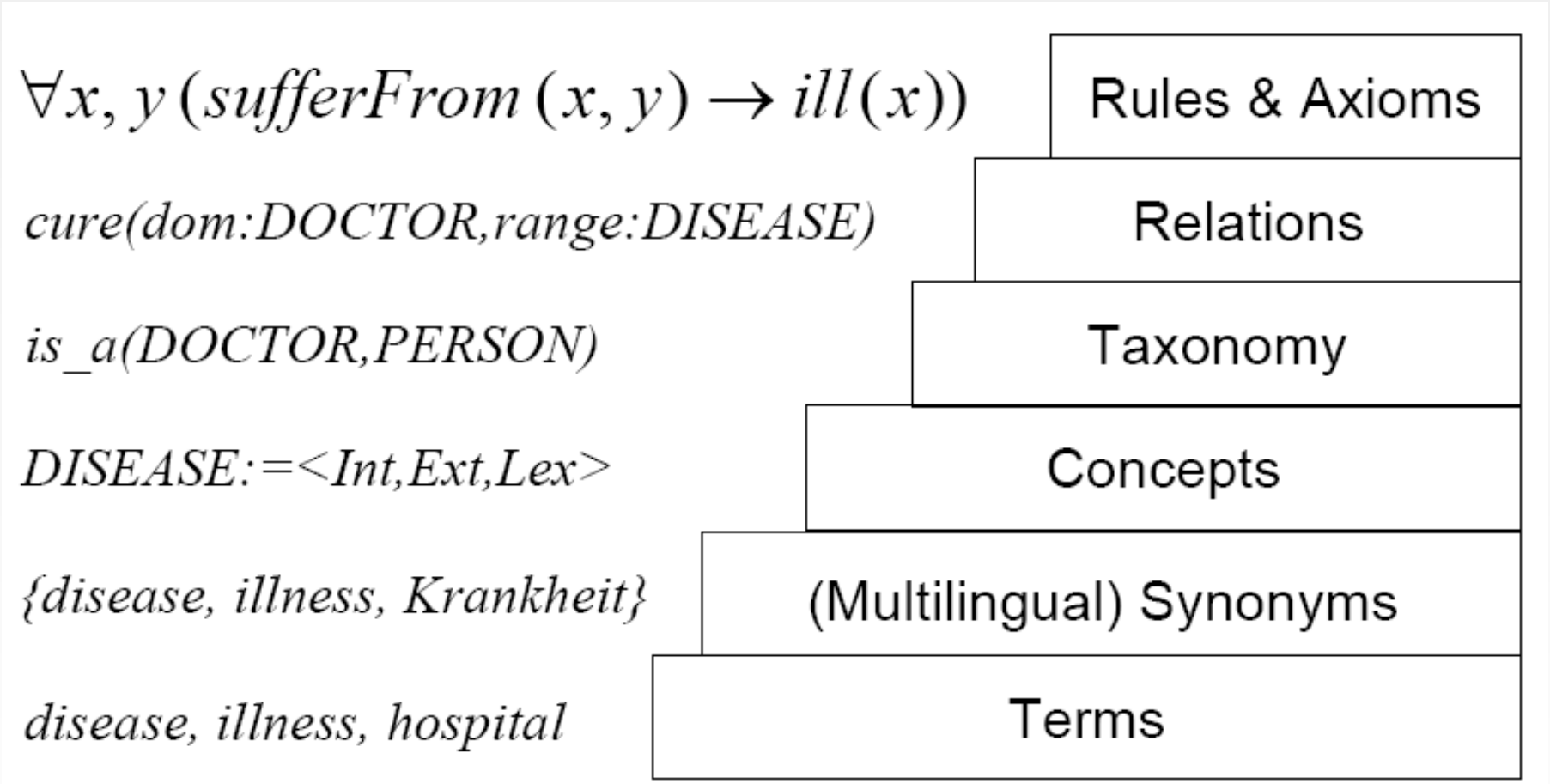
Why Text?

- Text is massively available on the Web
- Relevant texts contain relevant knowledge about a domain
- Linguistic knowledge remains associated with the ontology (Sintek et al., 2004)

OL as Reverse Engineering (Buitelaar et al., 2005)



OL from Text Layer Cake (Buitelaar et al., 2005)



Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Term Extraction - Linguistic Methods

- Part-of-speech tagging: Identify syntactic class
  - Ex: Noun -> Class, Verb -> Relation
- Stemming
  - Ex: **Formal**(ize/ization/ized/izing)
- Head-modifier analysis
  - Ex: Fast **car**, the hood of the **car**
- Grammatical function analysis
  - Ex: “John played football in the garden” -> play(John,football)

Term Extraction - Other methods

- Statistical Methods
  - Term Weighting (TF-IDF)
  - Co-occurrence analysis (Common method applied in Text Mining)
  - Comparison of frequencies between domain and general corpora
- Hybrid Methods
  - Linguistic rules to extract term candidates
  - Statistical (pre- or post-) filtering

Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Synonym Extraction

- Extending WordNet (Term Classification)
- Co-occurrence between terms (Term Clustering)

Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Concept Extraction

A term may indicate a concept, if we define its:

- Intension
  - (In)formal definition of the objects this concept describes
  - Ex: A **disease** is an impairment of health or a condition of abnormal functioning
- Extension
  - Set of objects described by this concept
  - Ex: Cancer, heart disease
- Lexical Realizations
  - The term itself and its multilingual synonyms

Intension

- Informal definition - a shallow definition as used in WordNet
  - Find the appropriate WordNet concept for a term and the appropiate conceptual relations (Navigli and Velardi, 2004)
- Formal definition - formal constraints defining class membership
  - Formal Concept Analysis

Extension

- Extraction of instances for a concept from text (Ontology Population)
- Relates to Knowledge Markup and Tag Suggestion (Semantic Metadata)
- Use Named-Entity Recognition
  - Ex: John is a football player -> **John** (Person) is an instance of **Football Player**
- Instances can be:
  - Names for objects
    - Ex: Person, Organization, Country, City
  - Event instances
    - Ex: Football Match (with Teams, Players, Officials, etc)

Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

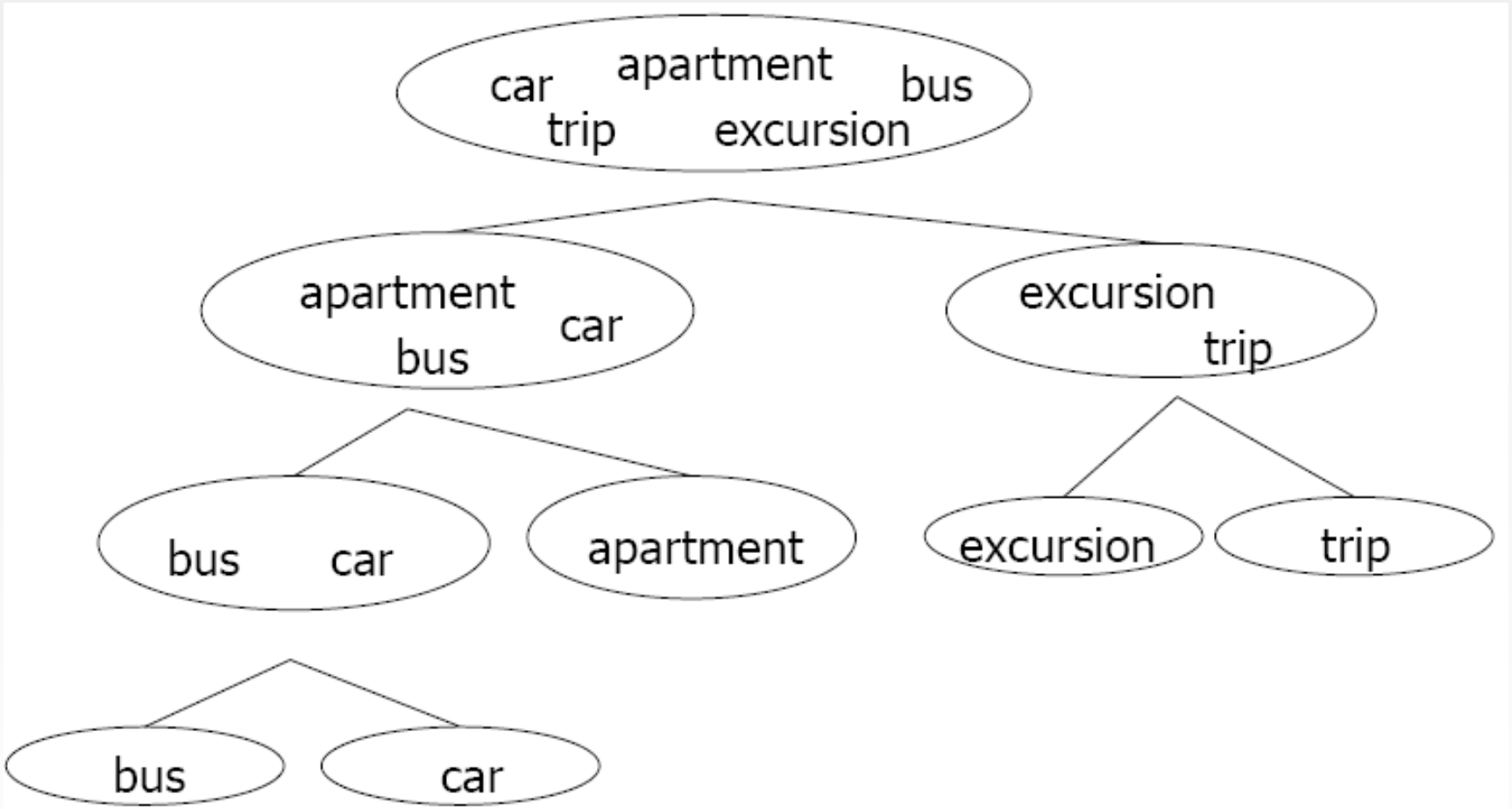
Taxonomy Extraction

- Lexico-syntactic patterns
- Clustering
- Linguistic approaches
- Document subsumption
- Combinations and other methods

Hearst Patterns (Hearst, 1992)

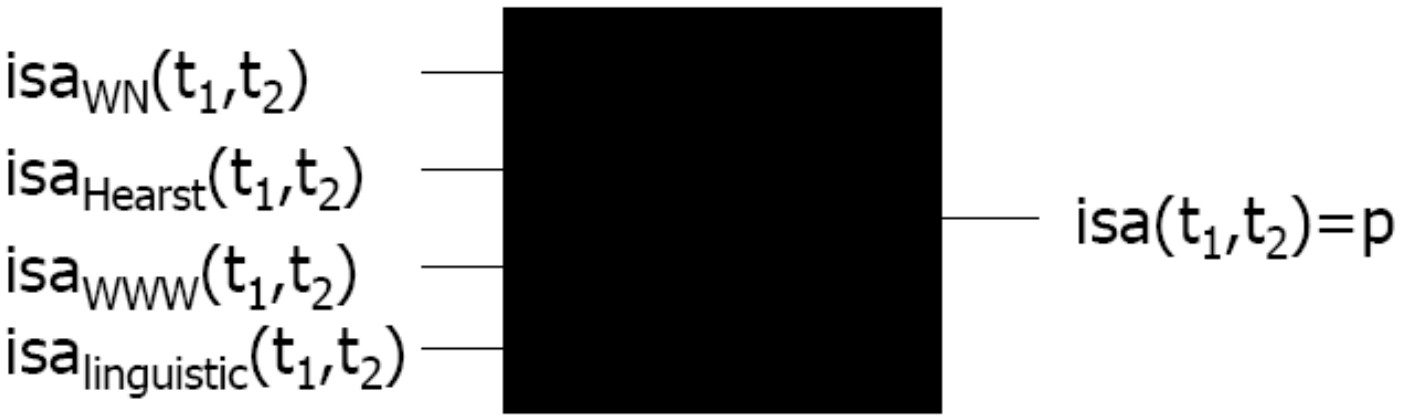
- Vehicles **such as** cars, trucks and bikes
- **Such** fruits **as** oranges or apples
- Swimming, running **and other** activities
- Publications, **especially** papers and books
- A salmon **is a** fish (Concept X Taxonomy Extraction)

Hierarchical Clustering



Other methods

- **Linguistic approach** - Use of modifiers (Navigli and Velardi, 2004; Buitelaar et al., 2004; Maedche and Staab, 2001)
  - **isa**(international credit card, credit card)
- **Document subsumption** - Term t_1 subsumes term t_2 [**is-a**(t_2,t_1)] if t_1 appears in all the documents in which t_2 appears
- **Combination method** - Tries to find an optimal combination of techniques using supervised ML



Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Relation Extraction - Specific Relations

- X consists of Y (**part-of**)
  - **The framework for OL** consists of **information extraction, ontology discovery and ontology organization**
- X **is used for** Y (purpose)
  - **OL** is used for **OE**
- X **leads to** Y (causation)
  - **Good OL methods** lead to **good OE**
- **the** X **of** Y (attribute)
  - **The hood** of **the car** is red

General Relations

- OntoLT: Mapping rules (Buitelaar et al., 2004)
  - SubjToClass_PredToSlot
- TextToOnto (Maedche and Staab, 2001)
  - $love(man, woman) \wedge$
    $love(kid, mother) \wedge$
    $love(kid, grandfather) \Longrightarrow love(person, person)$
- Still, different verbs can represent the same (or a similar) relation
  - Clustering -> {advise, teach, instruct}

Subsubsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Rule Extraction

- DIRT - Discovery of Inference Rules from Text (Lin and Pantel, 2001)
  - Let X be an algorithm which solves a problem Y
  - Using similar constructions like **X solves Y**, **Y is solved by X**, **X resolves Y**
  - $\forall x, y\ solves(X, Y) \Rightarrow isSolvedBy(Y, X)$ (Inverse object property)
  - $\forall x, y\ solves(X, Y) \Rightarrow resolves(X, Y)$ (Equivalent object property)

Axiom Extraction

- Automated Evaluation of ONtologies - AEON (Völker et al., 2008)
  - Axioms are extracted (using lexico-syntatic patterns) from a Web Corpus
- Dealing with uncertainty and inconsistency (Haase and Völker, 2005)
  - Disjointness axioms -> disjoint(man,woman)
- These methods are important because text contains inconsistency

Example of OL from text: OntoLT (Buitelaar et al., 2004)

- Use of mapping rules
  - The predicate of a sentence is a **relation** or **slot**
- Mapping rules have corresponding operators **SubjToClass -> CreateCls()**
- Users validate classes and slots candidates

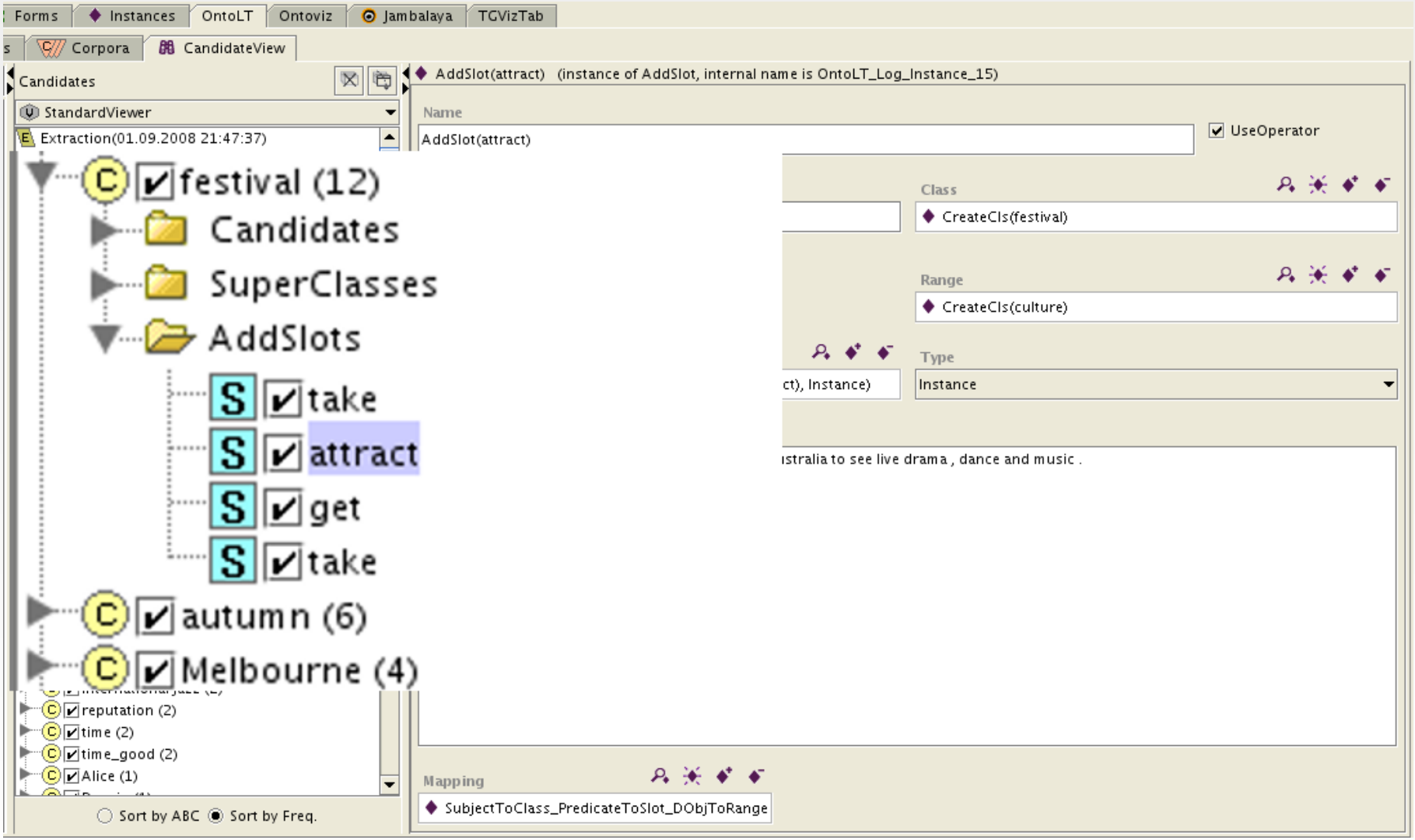
OntoLT

- Using sentences like

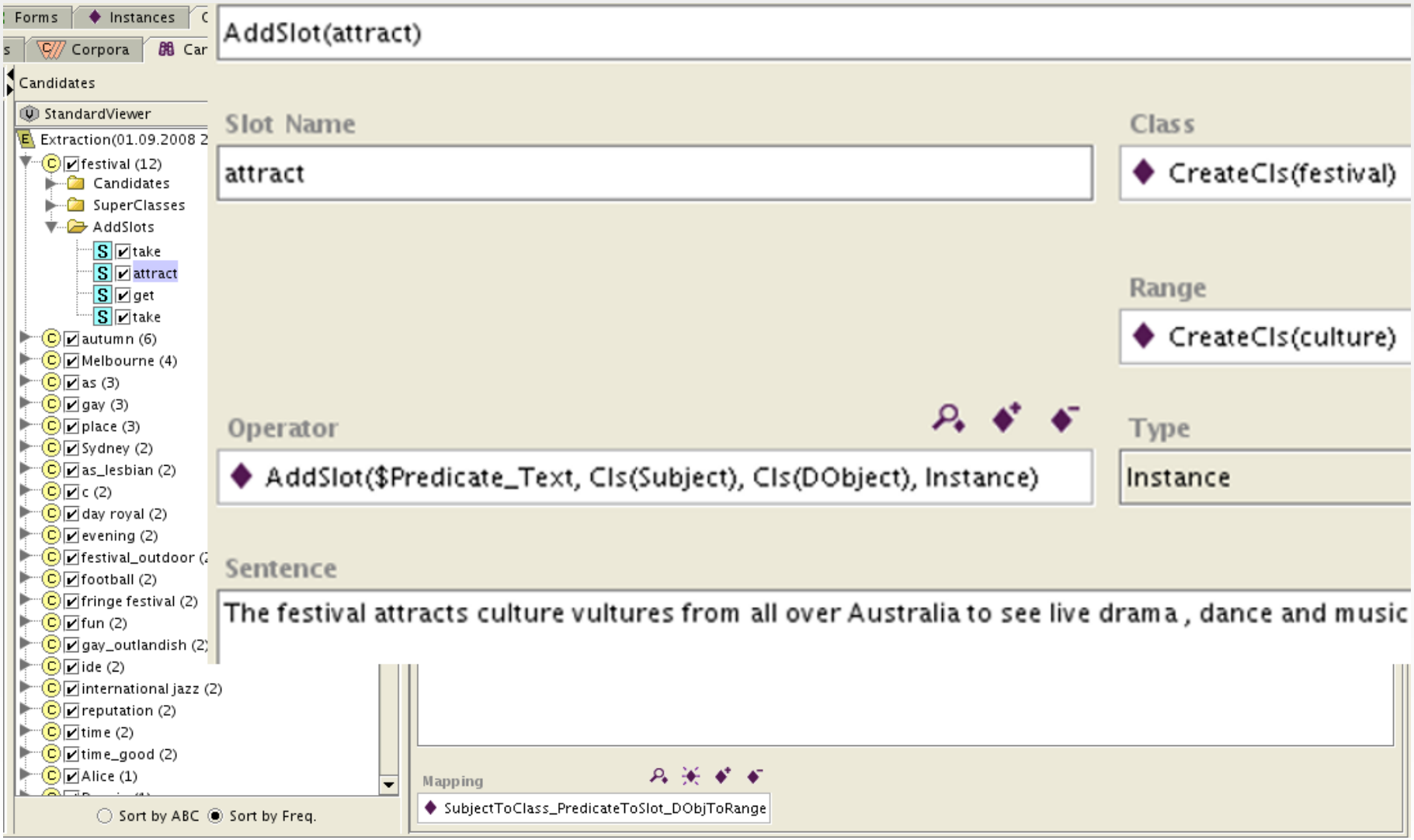
  The festival attracts culture vultures from all over Australia to see live drama, dance and music

  the system infers:
- **festival** and **culture** are class candidates - using statistical analysis (TF-IDF)
- **attracts** is a relation between **festival** and **culture** - using NLP

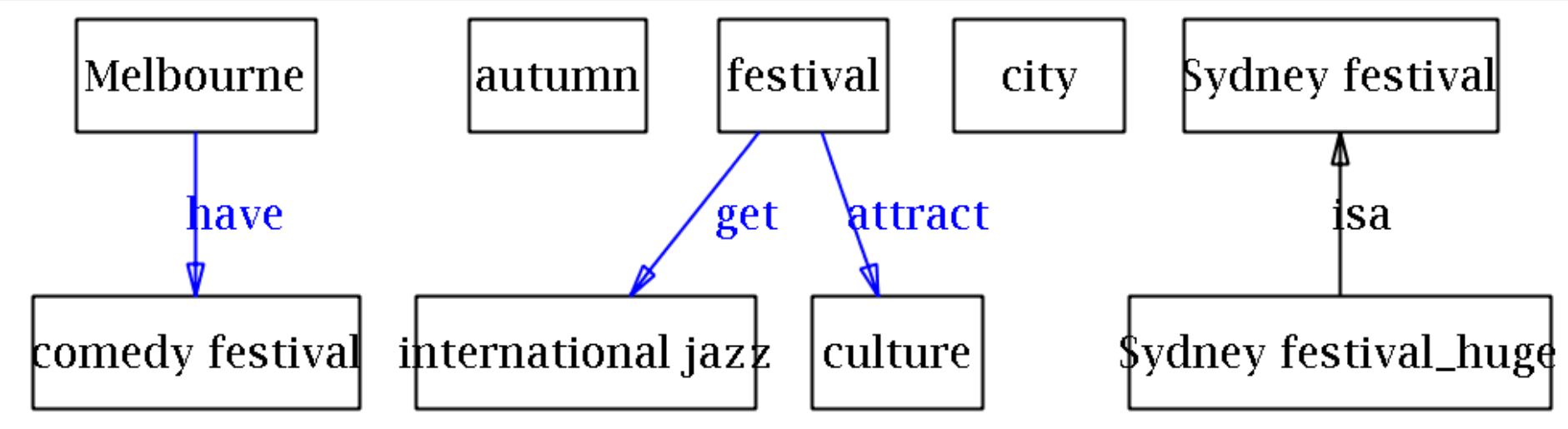
OntoLT Screenshot #1



OntoLT Screenshot #2



OntoLT: Extracted Ontology



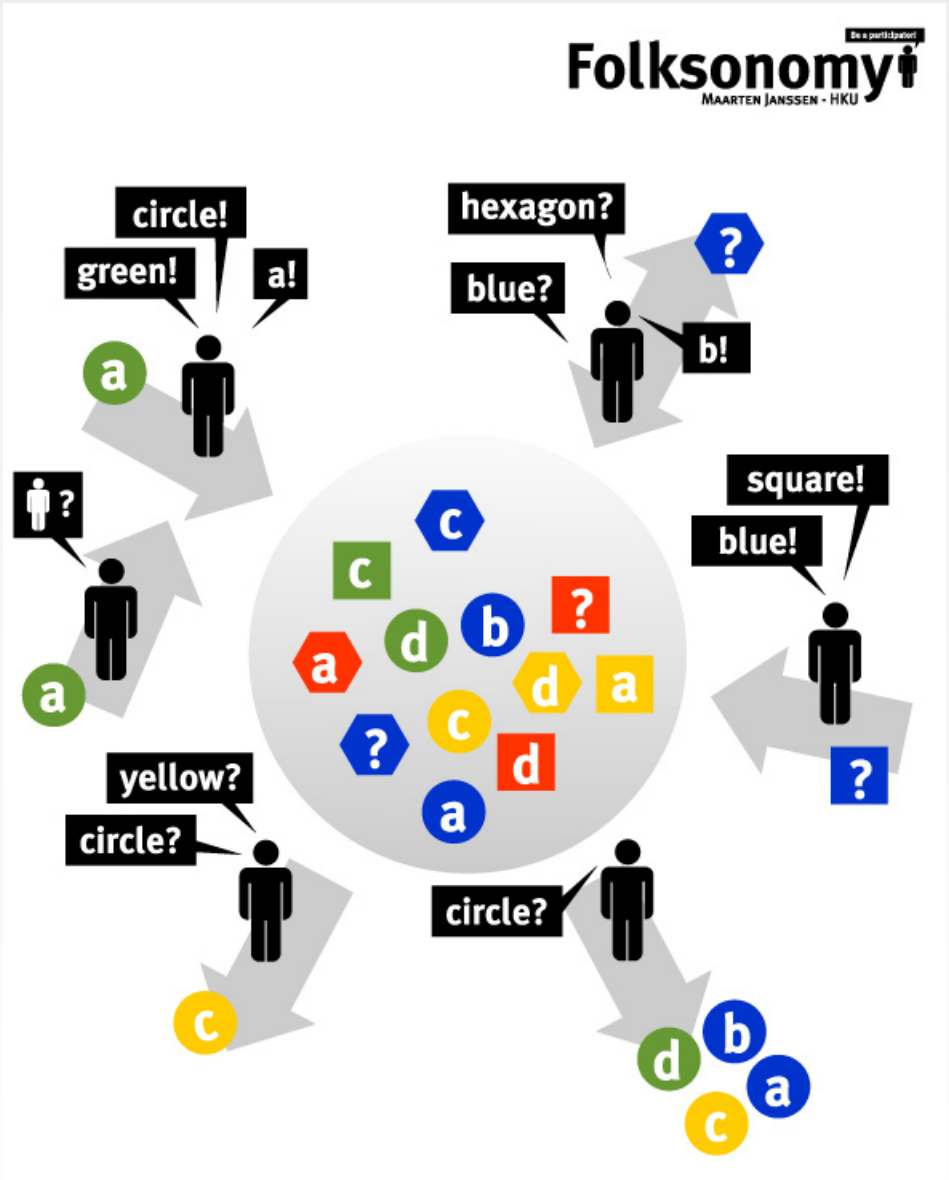
Subsections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

Folksonomies? Not yet!

Tag Cloud (Wikipedia, 2008b)

THIS is a Folksonomy (Pick, 2006)



Formal Definition of Folksonomy (Mika, 2007)

- Graph with hyper edges containing:
- $A = \{a_1, ..., a_k\}$ (Actors)
- $C = \{c_1, ..., c_l\}$ (Concepts)
- $I = \{i_1, ..., i_m\}$ (Instance of Objects - Web Resources)
- $T \subseteq A \times C \times I$ (Tags - Folksonomy)
- Two graphs: O_{ac} and O_{ci}

What does this have to do with OL? (Mika, 2007)

- Extract subsumption relations using set theory
- In O_{ci}, A is a superconcept of B if:
- The set of items classified under B is a subset of the entities under A
- $B \subseteq A \Leftrightarrow A \cap B = B$
- Overlapping set of instances (similar to document subsumption)

Concept Clustering Mika (2007)

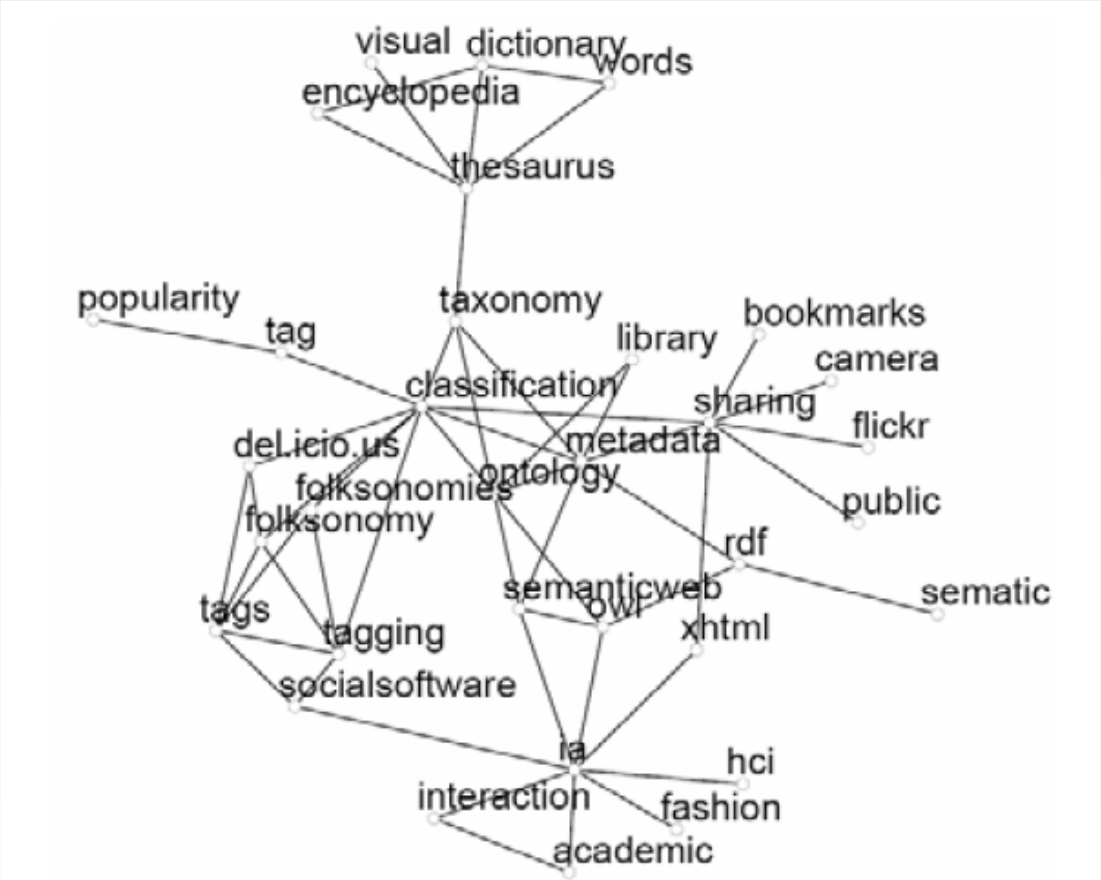
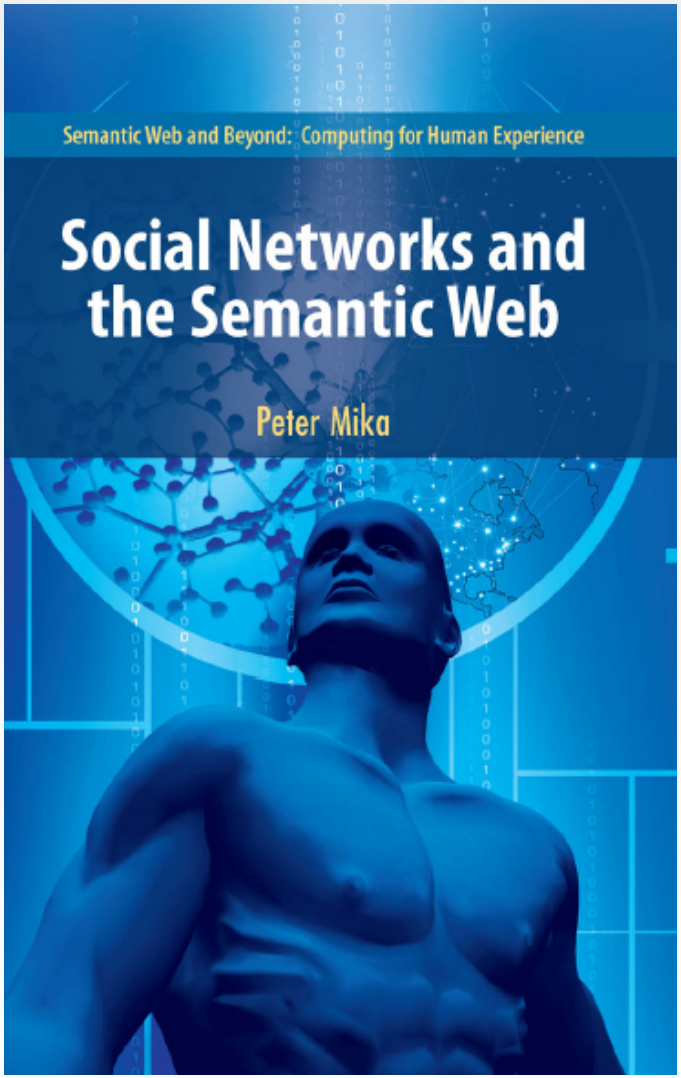
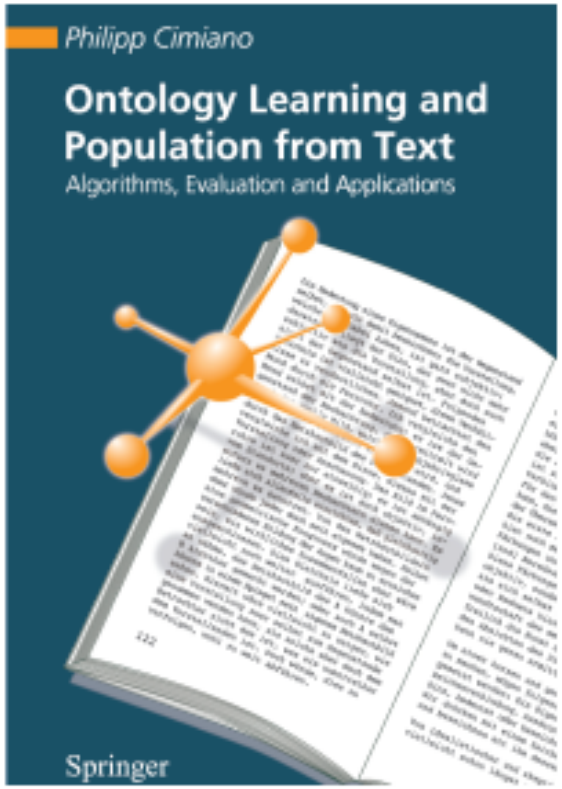


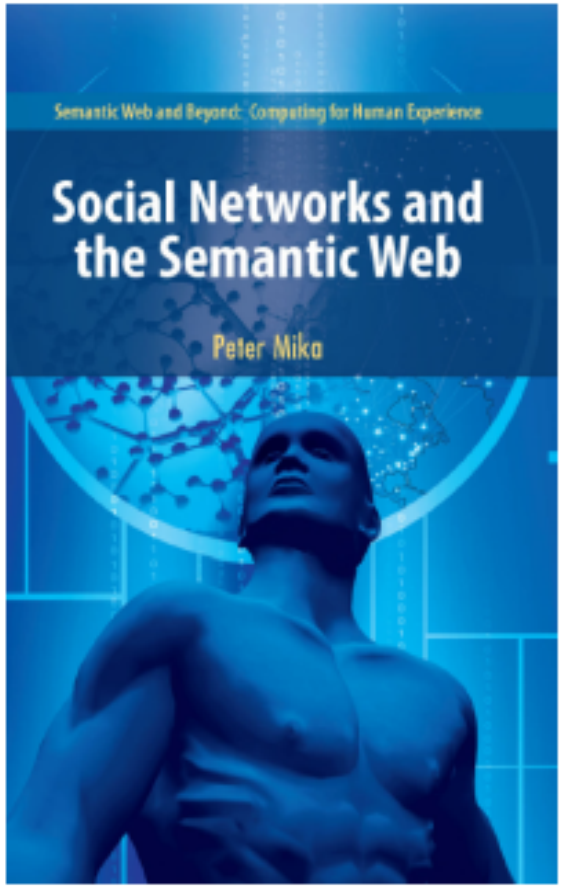
Figure: Del.icio.us tags: a 3-neighborhood of the term ontology (O_{ci})

OL from Social Network Analysis




Philipp Cimiano
Ontology Learning and Population from Text
Algorithms, Evaluation and Applications
Springer

+


Semantic Web and Beyond: Computing for Human Experience
Social Networks and the Semantic Web
Peter Mika

To appear!

Sections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

OL Tools

- ASIUM - Acquisition of SemantIc knowledge Using ML Methods (Faure and Edellec, 1998)
  - Taxonomic relations among terms in technical texts
  - Conceptual Clustering
- OntoLearn (Velardi et al., 2002)
  - Enrich a domain ontology with concepts and relations
  - NLP and ML

More OL Tools

- Text-To-Onto (Maedche and Volz, 2001)
  - Find taxonomic and non-taxonomic relations
  - Statistics, Pruning Techniques and Association Rules
  - Sucessor: OntoWare.org Text2Onto -> (Cimiano and Völker, 2005)
- OntoWare.org LExO - Learning Expressive Ontologies (Völker et al., 2007)
  - Transform natural language definitions into OWL DL axioms
- OntoLP - Engenharia de Ontologias em Língua Portuguesa (SBC2008)

Sections

1. Introduction
2. Methods
   - Ontology Learning from Text
     - Terms
     - Synonyms
     - Concepts
     - Taxonomy
     - Relations
     - Rules and Axioms
   - Ontology Learning from Folksonomies
3. Tools
4. Conclusion

How to evaluate OL?

- Non-formal methods
- 1st step: Formalize the task of OL from text (Sintek et al., 2004)
- Next steps:
  - Benchmark corpora and ontologies
  - Evaluation of methods using different information sources

The future

- We need ontologies!
- We need to build them quickly, easily and they have to be reliable!
  - Time: OL makes OE faster
  - Difficulty: OL makes OE easier
  - Confidence: Relevant text (like technical reports written by domain experts) are confident sources of information

References I

Buitelaar, P., Cimiano, P., Grobelnik, M., and Sintek, M. (2005). Ontology learning from text. Tutorial at ECML/PKDD 2005. Workshop on Knowledge Discovery and Ontologies. Porto, Portugal. http://www.aifb.uni-karlsruhe.de/WBS/pci/OL_Tutorial_ECML_PKDD_05/ECML-OntologyLearningTutorial-20050923.pdf.

Buitelaar, P., Olejnik, D., and Sintek, M. (2004). A protégé plug-in for ontology extraction from text based on linguistic analysis. In Bussler, C., Davies, J., Fensel, D., and Studer, R., editors, *ESWS*, volume 3053 of *Lecture Notes in Computer Science*, pages 31–44. Springer.

Cimiano, P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Cimiano, P. and Völker, J. (2005). Text2onto - a framework for ontology learning and data-driven change discovery. In Montoyo, A., Munoz, R., and Metais, E., editors, *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 3513 of *Lecture Notes in Computer Science*, pages 227–238, Alicante, Spain. Springer.

Faure, D. and Edellec, C. N. (1998). A corpus-based conceptual clustering method for verb frames and ontology acquisition. In *In LREC workshop on*, pages 5–12.

Haase, P. and Völker, J. (2005). Ontology learning and reasoning - dealing with uncertainty and inconsistency. In *In Proceedings of the Workshop on Uncertainty Reasoning for the Semantic Web (URSW*, pages 45–55.

References II

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *In Proceedings of the 14th International Conference on Computational Linguistics*, pages 539–545.

Lin, D. and Pantel, P. (2001). Dirt @sbt@discovery of inference rules from text. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 323–328, New York, NY, USA. ACM.

Maedche, A. and Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2):72–79.

Maedche, E. and Volz, R. (2001). The ontology extraction and maintenance framework text-to-onto. In *In Proceedings of the ICDM'01 Workshop on Integrating Data Mining and Knowledge Management*.

Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1):5–15.

Navigli, R. and Velardi, P. (2004). Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics*, 30(2):151–179.

OntoSum (2008). Ontology learning. `http://www.ontosum.org/?q=node/17`. [Online; accessed 31-August-2008].

Pick, M. (2006). Social bookmarking services and tools: The wisdom of crowds that organizes the web - robin good's latest news##.

References III

Sintek, M., Buitelaar, P., and Olejnik, D. (2004). A formalization of ontology learning from text. In *Proc. of the Workshop on Evaluation of Ontology-based Tools (EON2004) at the International Semantic Web Conference*.

Velardi, P., Navigli, R., and Missikoff, M. (2002). An integrated approach for web ontology learning and engineering. *IEEE Computer*.

Völker, J., Vrandečić, D., Sure, Y., and Hotho, A. (2008). Aeon - an approach to the automatic evaluation of ontologies. *Appl. Ontol.*, 3(1-2):41–62.

Völker, J., Hitzler, P., and Cimiano, P. (2007). Acquisition of owl dl axioms from lexical resources. In Franconi, E., Kifer, M., and May, W., editors, *Proceedings of the 4th European Semantic Web Conference (ESWC'07)*, volume 4519 of *Lecture Notes in Computer Science*, pages 670–685. Springer.

Wikipedia (2008a). Ontology learning — wikipedia, the free encyclopedia. [Online; accessed 31-August-2008].

Wikipedia (2008b). Tag cloud — wikipedia, the free encyclopedia. [Online; accessed 10-September-2008].