# Semantic Web and Machine Learning Tutorial

**Steffen Staab**
*ISWeb – Information Systems and Semantic Web*
**University of Koblenz**
**Germany**

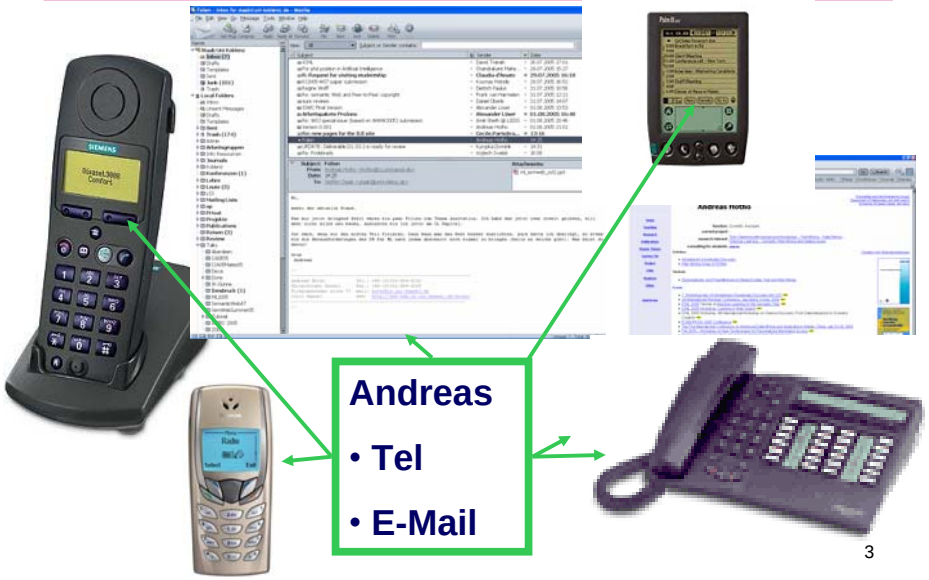
**Andreas Hotho**
**Knowledge and Data Engineering Group**
**University of Kassel**
**Germany**

## Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications



## Syntax is not enough



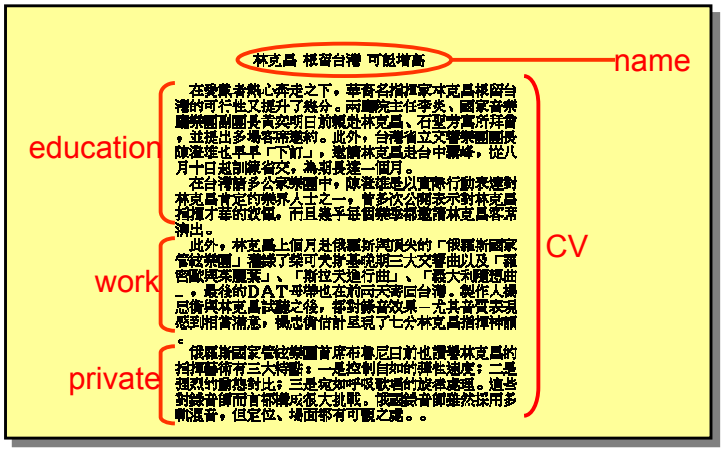



## Information Convergence

- Convergence not just in devices, also in "information"
  - Your personal information (phone, PDA,...)
    Calendar, photo, home page, files...
  - Your "professional" life (laptop, desktop, ... Grid)
    Web site, publications, files, databases, ...
  - Your "community" contexts (Web)
    Hobbies, blogs, fanfic, social networks...
- The Web teaches us that people will work to share
  - How do we CREATE, SEARCH, and BROWSE in the non-text based parts of our lives?

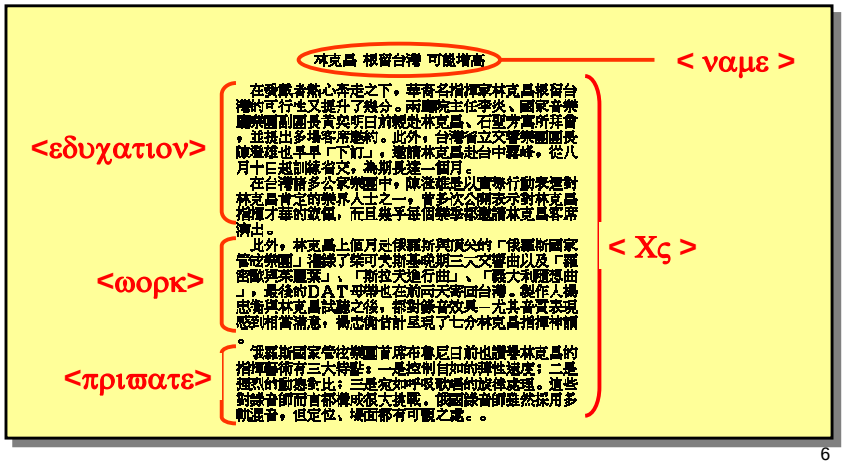
# Meaning of Informationen:

**(or: what it means to be a computer)**






# XML ≠ Meaning, XML = Structure






# Source of Problems

XML is unspecific:

❶ No predetermined vocabulary

❷ No semantics for relationships

⇨ ❶ & ❷ must be specified upfront

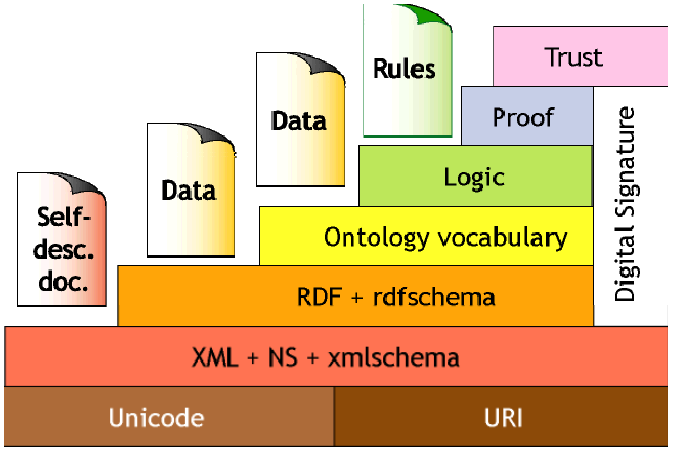
Only possible in close cooperations

- Small, reasonably stable group ☹
- Common interests or authorities

Not possible in the Web or on a broad scale in general !



# (One) Layer Model of the Semantic Web



# Some Principal Ideas

- URI – uniform resource identifiers
- XML – common syntax
- Interlinked
- Layers of semantics – from database to knowledge base to proofs

Tim Berners-Lee, Weaving the Web

***Design principles of WWW applied to Semantics!!***



# What is an Ontology?

Gruber 93:

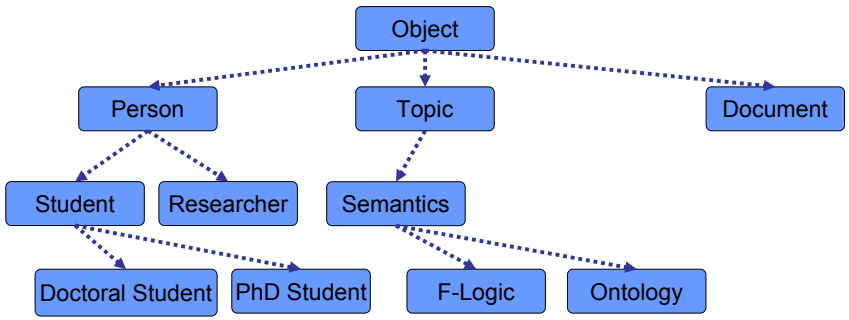
An Ontology is a

formal specification ⇒ Executable
of a shared ⇒ Group of persons
conceptualization ⇒ About concepts
of a domain of interest ⇒ Between application and „unique truth“





## Taxonomy

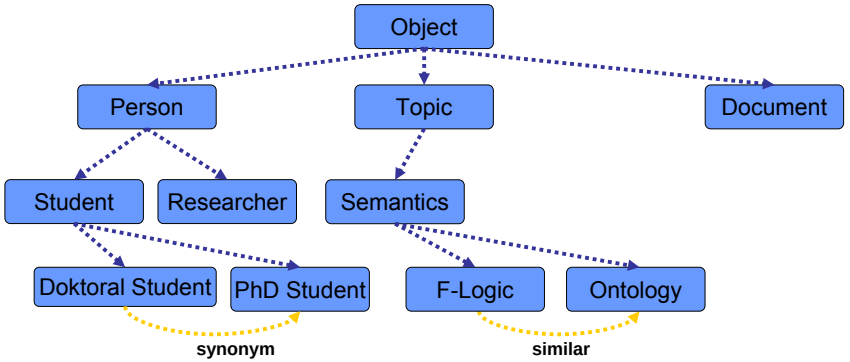



**Taxonomy := Segmentation, classification and ordering of elements into a classification system according to their relationships between each other**



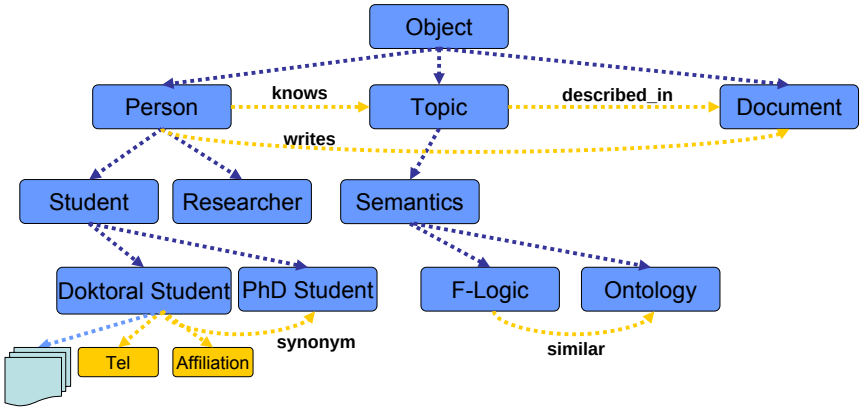


## Thesaurus




- **Terminology for specific domain**
- **Graph with primitives, 2 fixed relationships (similar, synonym)**
- **originate from bibliography**

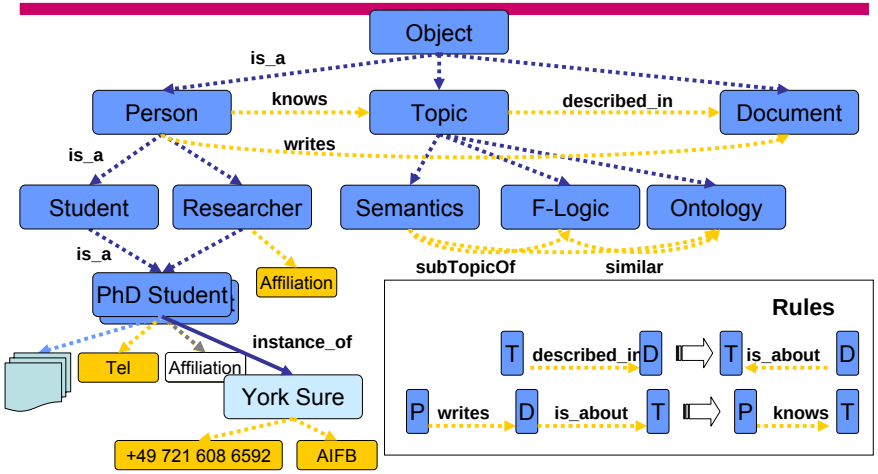
## Topic Map




- **Topics (nodes), relationships and o*ccurences* (to documents)**
- **ISO-Standard**
- **typically for navigation- and visualisation**



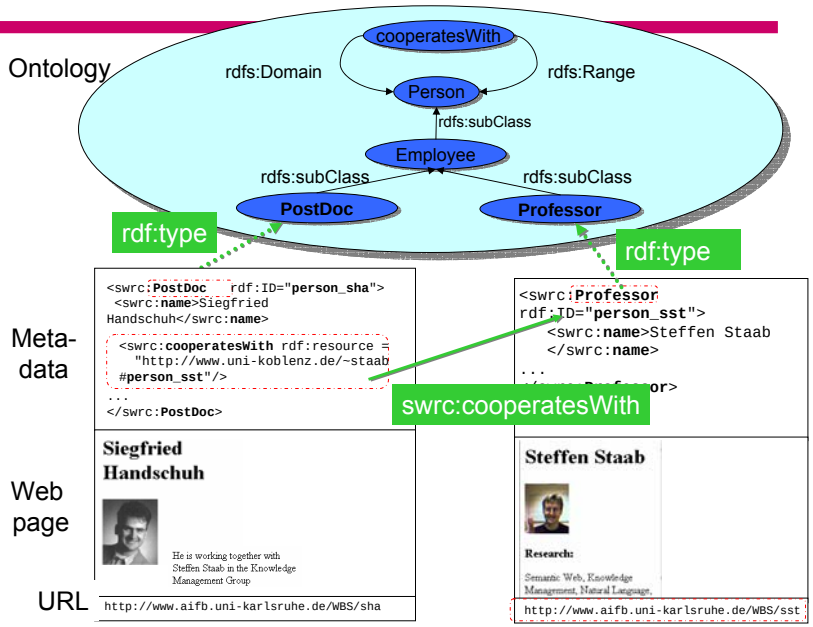
## Ontology (in our sense)




- **Representation Language: Predicate Logic (F-Logic)**
- **Standards: RDF(S); coming up standard: OWL**



## The Semantic Web




## What's in a link? Formally

W3C recommendations
- RDF: an edge in a graph
- OWL: consistency (+subsumption+classif. + ...)

Currently under discussion
- Rules: a deductive database

Currently under intense research
- Proof: worked-out proofs
- Trust: signature & everything working together

## What's in a link? Informally

- RDF: pointing to shared data
- OWL: shared terminology
- Rules: if-then-else conditions
- Proof: proof already shown
- Trust: reliability

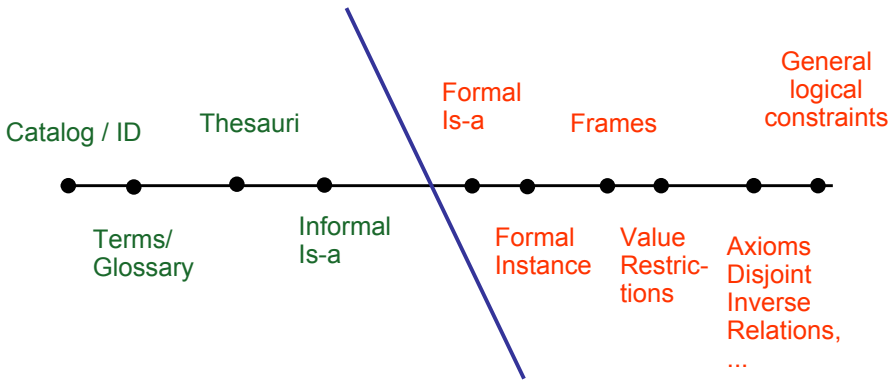


## Ontologies and their Relatives (I)

- There are many relatives around:
  - **Controlled vocabularies, thesauri and classification systems available in the WWW,** see http://www.lub.lu.se/metadata/subject-help.html
    - Classification Systems (e.g. UNSPSC, Library Science, etc.)
    - Thesauri (e.g. Art & Architecture, Agrovoc, etc.)
    - DMOZ Open Directory http://www.dmoz.org
  - **Lexical Semantic Nets**
    - WordNet, see http://www.cogsci.princeton.edu/~wn/
    - EuroWordNet, see http://www.hum.uva.nl/~ewn/
  - **Topic Maps, http://www.topicmaps.org** (e.g. used within knowledge management applications)
- In general it is difficult to find the border line!



## Ontologies and their Relatives (II)

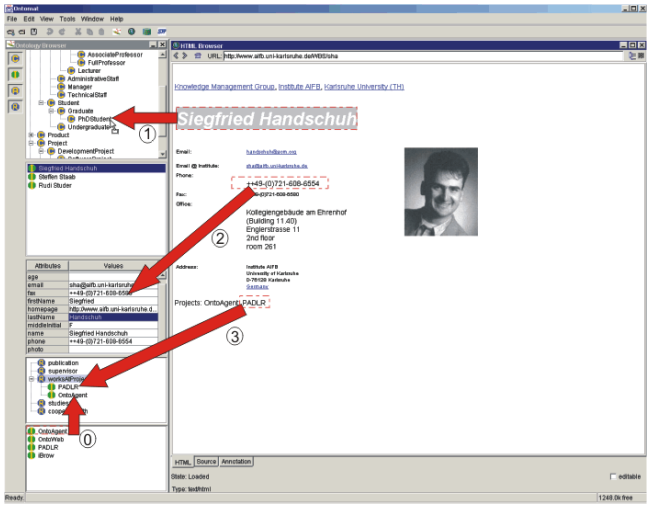





## Ontologies - Some Examples

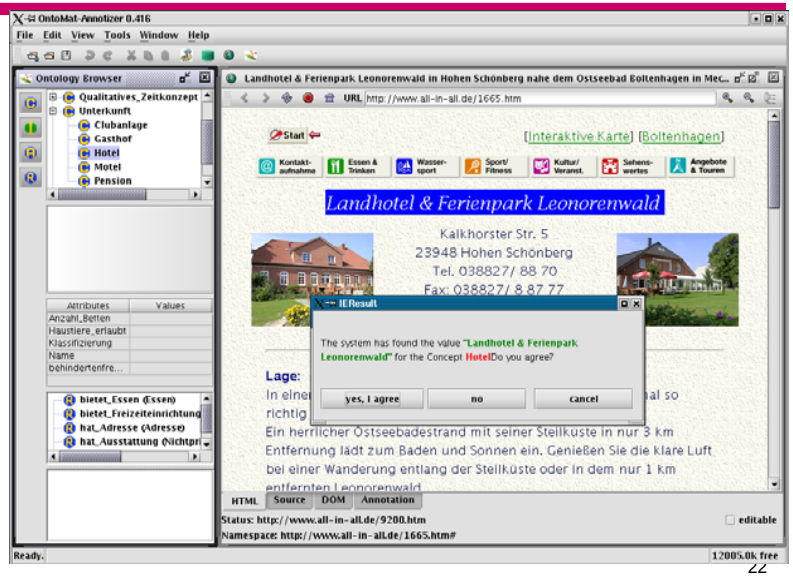
- General purpose ontologies:
  - WordNet / EuroWordNet, http://www.cogsci.princeton.edu/~wn
  - The Upper Cyc Ontology, http://www.cyc.com/cyc-2-1/index.html
  - IEEE Standard Upper Ontology, http://suo.ieee.org/
- Domain and application-specific ontologies:
  - RDF Site Summary RSS, http://groups.yahoo.com/group/rss-dev/files/schema.rdf
  - UMLS, http://www.nlm.nih.gov/research/umls/
  - GALEN
  - SWRC – Semantic Web Research Community: http://ontoware.org/projects/swrc/
  - RETSINA Calendering Agent, http://ilrt.org/discovery/2001/06/schemas/ical-full/hybrid.rdf
  - Dublin Core, http://dublincore.org/
- Web Services Ontologies
  - Core ontology of services http://cos.ontoware.org
  - Web Service Modeling ontology http://www.wsmo.org
  - DAML-S
- Meta-Ontologies
  - Semantic Translation, http://www.ecimf.org/contrib/onto/ST/index.html
  - RDFT, http://www.cs.vu.nl/~borys/RDFT/0.27/RDFT.rdfs
  - Evolution Ontology, http://kaon.semanticweb.org/examples/Evolution.rdfs
- Ontologies in a wider sense
  - Agrovoc, http://www.fao.org/agrovoc/
  - Art and Architecture, http://www.getty.edu/research/tools/vocabulary/aat/
  - UNSPSC, http://eccma.org/unspsc/
  - DTD standardizations, e.g. HR-XML, http://www.hr-xml.org/

# Tools for markup...



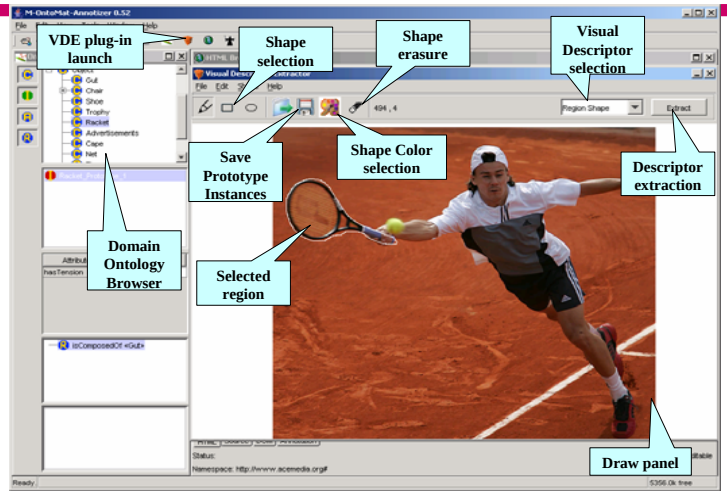



# Not tied to specific domains





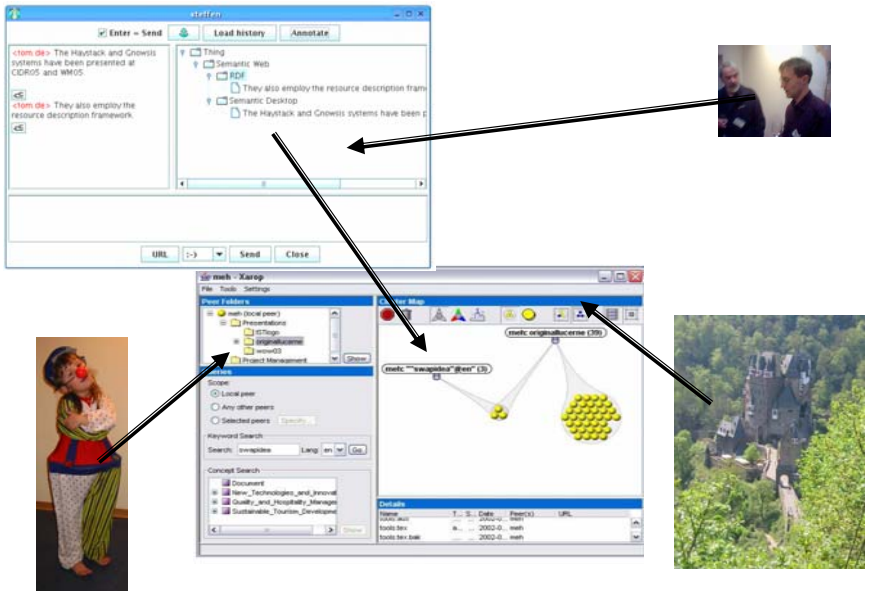

# Not tied to specific domains



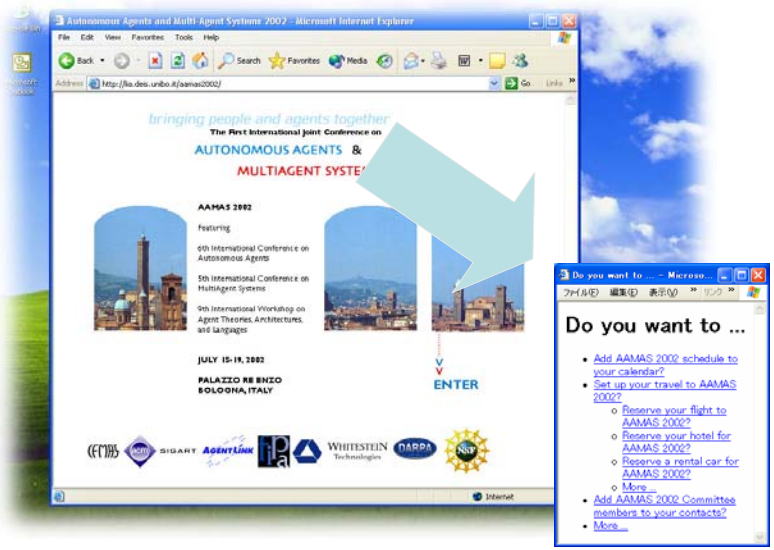

M-OntoMat is publicly available
http://acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html



# Shared Workspace (Xarop + Screenshot)



# Coming sooner than you may think…






# Social networks:
## *e.g.* Friend of a Friend (FOAF)

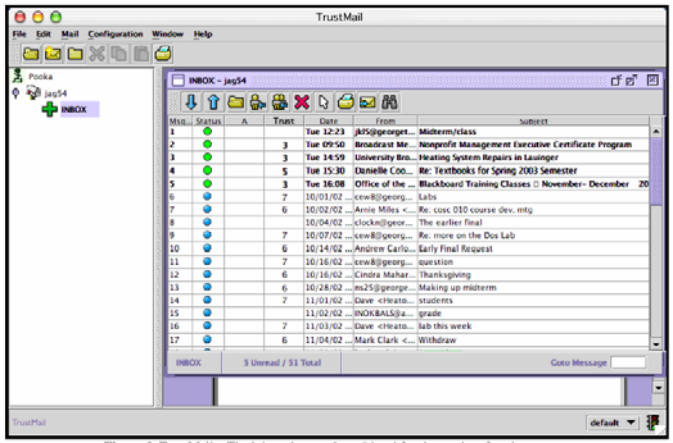
- Say stuff about yourself (or others) in OWL files, link to who you "know"

```
- <rdf:RDF>
  - <foaf:PersonalProfileDocument rdf:about="">
      <foaf:maker rdf:nodeID="me"/>
      <foaf:primaryTopic rdf:nodeID="me"/>
      <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
      <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
    </foaf:PersonalProfileDocument>
  - <foaf:Person rdf:nodeID="me">
      <foaf:name>Steffen Staab</foaf:name>
      <foaf:title>Prof. Dr.</foaf:title>
      <foaf:givenname>Steffen</foaf:givenname>
      <foaf:family_name>Staab</foaf:family_name>
      <foaf:mbox_sha1sum>ae8f32f31b69df2872d4e5d2e3bf21cf09cadf90</foaf:mbox_sha1sum>
      <foaf:homepage rdf:resource="http://www.uni-koblenz.de/~staab/"/>
      <foaf:depiction rdf:resource="http://www.uni-koblenz.de/~staab/images/Steffen7.jpg"/>
      <foaf:phone rdf:resource="tel:+49-261-2872761"/>
      <foaf:workplaceHomepage rdf:resource="http://isweb.uni-koblenz.de"/>
      <foaf:workInfoHomepage rdf:resource="Semantic%20Web,%20Research%20,%20Teaching"/>
      <foaf:schoolHomepage rdf:resource="http://www.jsg-karlstadt.de/"/>
      <foaf:projectHomepage rdf:resource="http://www.acemedia.org/"/>
      <foaf:projectHomepage rdf:resource="http://asg-platform.org/cgi-bin/twiki/view/Public/WebHome"/>
      <foaf:projectHomepage rdf:resource="http://www.projecthalo.com/"/>
      <foaf:projectHomepage rdf:resource="http://swap.semanticweb.org/"/>
      <foaf:projectHomepage rdf:resource="http://bibster.semanticweb.org/"/>
      <foaf:groupHomepage rdf:resource="http://isweb.uni-koblenz.de/"/>
      <foaf:interest rdf:resource="http://www.wissensmanagement-gesellschaft.de/"/>
```

Estir



# Using FOAF in other contexts

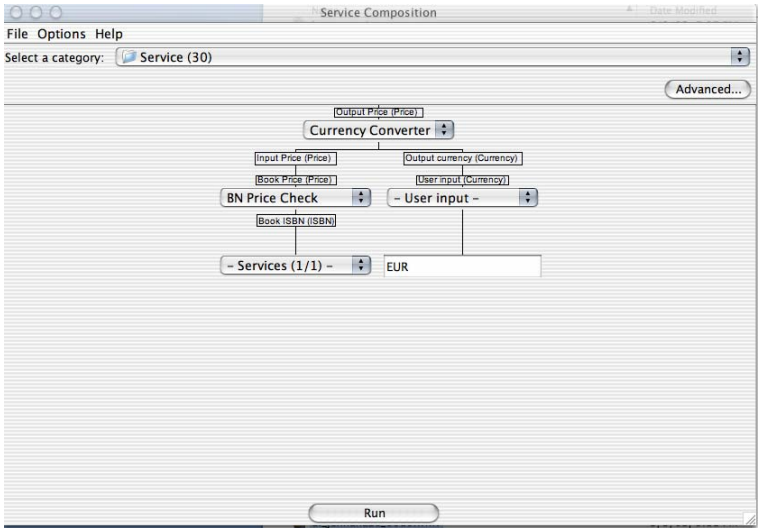



Jennifer Golbeck

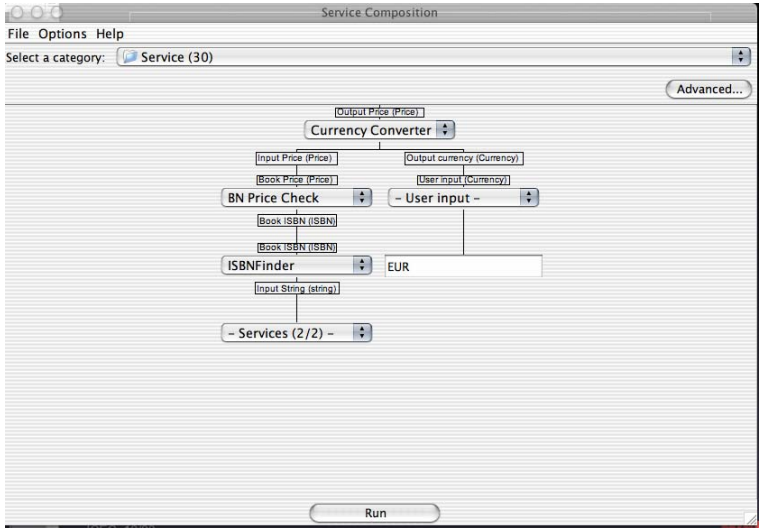
http://trust.mindswap.org



# Get a B&N price (In Euros)

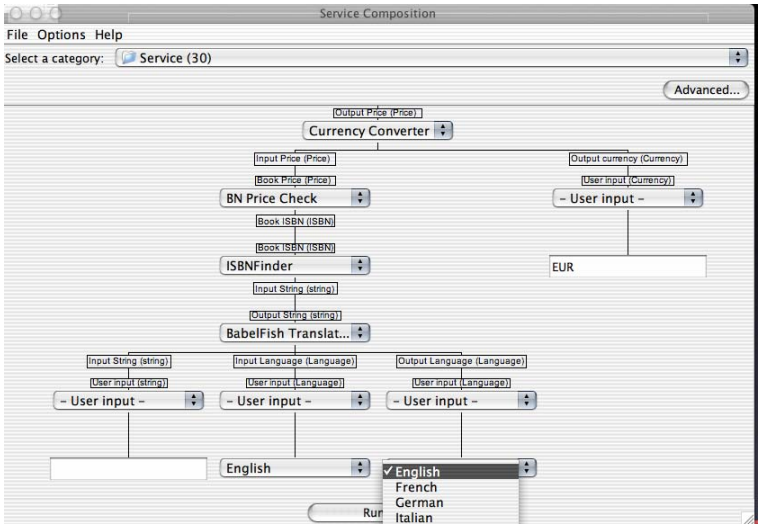


# Of a particular book

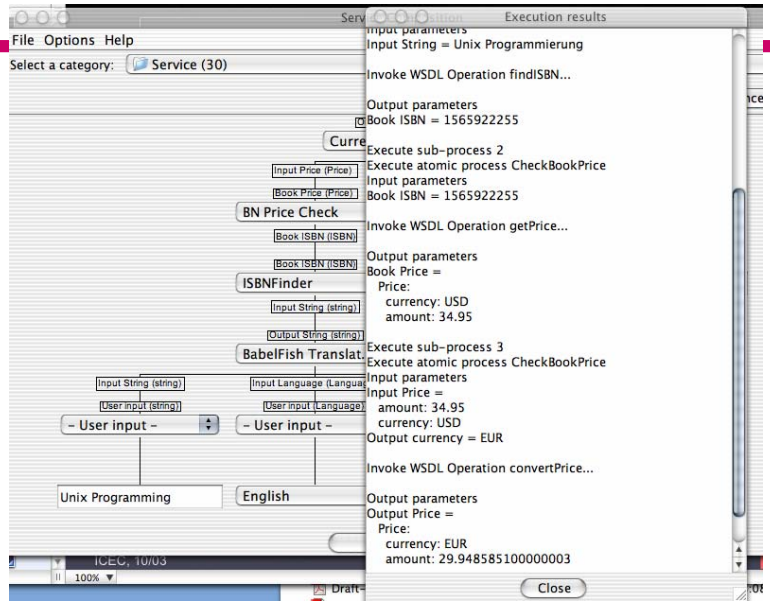





# In its German edition?

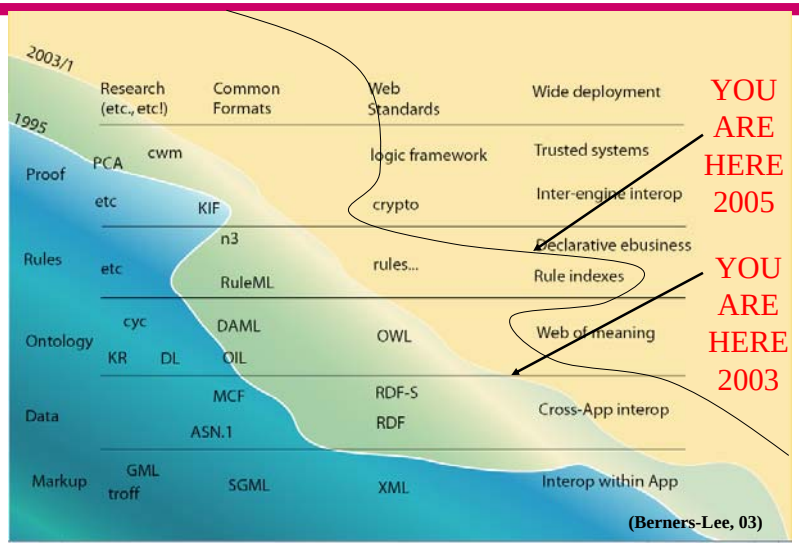







# The Semantic Wave



# Now.

- RDF, RDFS and OWL are ready for prime time
  - Designs are stable, implementations maturing
- Major Research investment translating into application development and commercial spinoffs
  - Adobe 6.0 embraces RDF
  - IBM releases tools, data and partnering
  - HP extending Jena to OWL
  - OWL Engines by Ontoprise GmbH, Network Inference, Racer GmbH
  - Proprietary OWL ontologies for vertical markets
    - c.f. pharmacology, HMO/health care, ... Soft drinks
  - Several new starts in SW space



# The semantic web and machine learning

What can machine learning do for the Semantic Web?

1. Learning Ontologies (even if not fully automatic)
2. Learning to map between ontologies
3. Deep Annotation: Reconciling databases and ontologies
4. Annotation by Information Extraction
5. Duplicate recognition

What can the Semantic Web do for Machine Learning?

1. Lots and lots of tools to describe and exchange data for later use by machine learning methods in a canonical way!
2. Using ontological structures to improve the machine learning task
3. Provide background knowledge to guide machine learning



# Foundations of the Semantic Web: References


- Semantic Web Activity at W3C http://www.w3.org/2001/sw/
- www.semanticweb.org (currently relaunched)
- Journal of Web Semantics
- D. Fensel et al.: Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press 2003
- G. Antoniou, F. van Harmelen. A Semantic Web Primer, MIT Press 2004.
- S. Staab, R. Studer (eds.). Handbook on Ontologies. Springer Verlag, 2004.
- S. Handschuh, S. Staab (eds.). Annotation for the Semantic Web. IOS Press, 2003.
- International Semantic Web Conference series, yearly since 2002, LNCS
- World Wide Web Conference series, ACM Press, first Semantic Web papers since 1999
- York Sure, Pascal Hitzler, Andreas Eberhart, Rudi Studer, The Semantic Web in One Day, *IEEE Intelligent Systems*, http://www.aifb.uni-karlsruhe.de/WBS/phi/pub/sw_inoneday.pdf

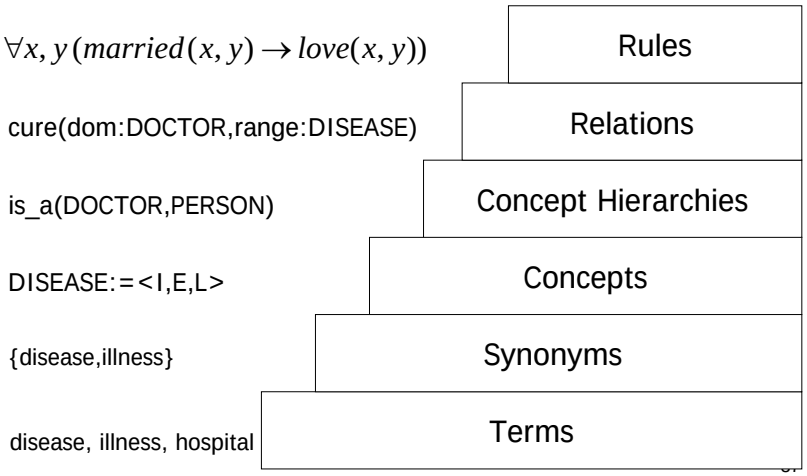

- Some slides have been stolen from various places, from Jim Hendler and Frank van Harmelen, in particular.



# Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications

# The OL Layer Cake




# How do people acquire taxonomic knowledge?

- I have no idea!
- But people apply taxonomic reasoning!
  - „Never do harm to any animal!“
  => „Don‘t do harm to the cat!“
- More difficult questions:
  - representation
  - reasoning patterns
- But let‘s speculate a bit! ;-)



# How do people acquire taxonomic knowledge?

What is liver cirrhosis?

Mr. Smith **died from** liver cirrhosis.
Mr. Jagger **suffers from** liver cirrhosis.
Alcohol abuse can **lead to** liver cirrhosis.

=>prob(isa(liver cirrhosis,disease))

# How do people acquire taxonomic knowledge?

What is liver cirrhosis?

Diseases **such as** liver cirrhosis are difficult to cure. (New York Times)

## How do people acquire taxonomic knowledge?

What is liver cirrhosis?

*Cirrhosis:* noun[uncountable] serious disease of the liver, often caused by drinking too much alcohol

$$\text{liver cirrhosis} \approx \text{cirrhosis} \wedge \text{isa}(\text{cirrhosis}, \text{disease})$$
$$\rightarrow \text{prob}(\text{isa}(\text{liver cirrhosis}, \text{disease}))$$



## Evaluation of Ontology Learning

The apriori approach is based on a gold standard ontology:

- Given an ontology modeled by an expert
  -> The so called gold standard
- Compare the learned ontology with the gold standard

- Which methods exists:
  - learning accuracy/precision/recall/f-measure
  - Count edges in the "ontology graph"
    - Counting of direct relation only (Reinberger et.al. 2005)
    - Least common superconcept
    - Semantic cotopy
    - ...
  - Evaluation via application (cf. section using ontologies)



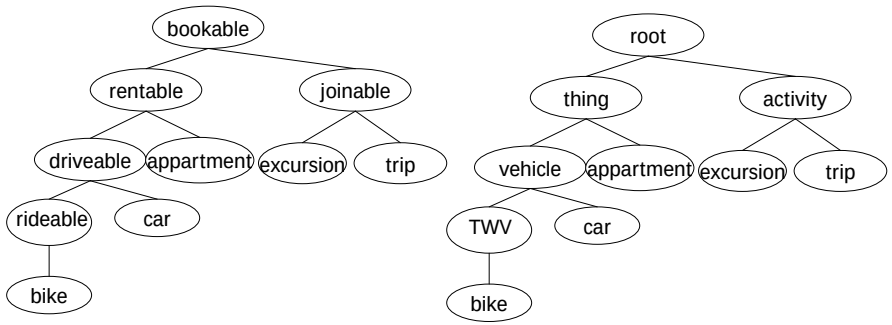
## The Semantic Cotopy

$$SC(c,O) = \{c' \mid c' \leq_O c \vee c \leq_O c'\}$$

**[Maedche & Staab 02]**



## Example for SC




SC(bike)={bike,rideable,driveable.rentable,bookable} SC(bike)={bike,TWV,vehicle,thing,root}

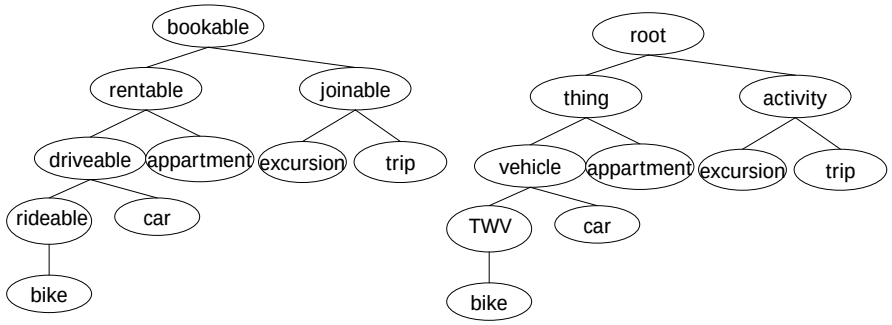
=> TO(bike,$O_1$,$O_2$)=1/9!!!

## Common Semantic Cotopy

$$SC'(c, O_1, O_2) = \{c' \mid c' \in C_1 \cap C_2 \wedge (c' \leq_{O_1} c \vee c \leq_{O_1} c')\}$$



## Example for SC‘

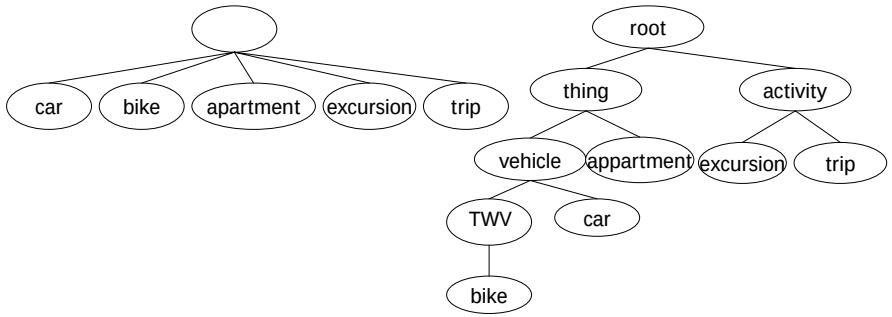



SC‘(driveable)={bike,car}  SC‘(vehicle)={bike,car}

=> TO(driveable,$O_1$,$O_2$)=1



## One more Example




SC‘(car)={car}  SC‘(vehicle)={bike,car}

=> TO(driveable,$O_1$,$O_2$)=1/2

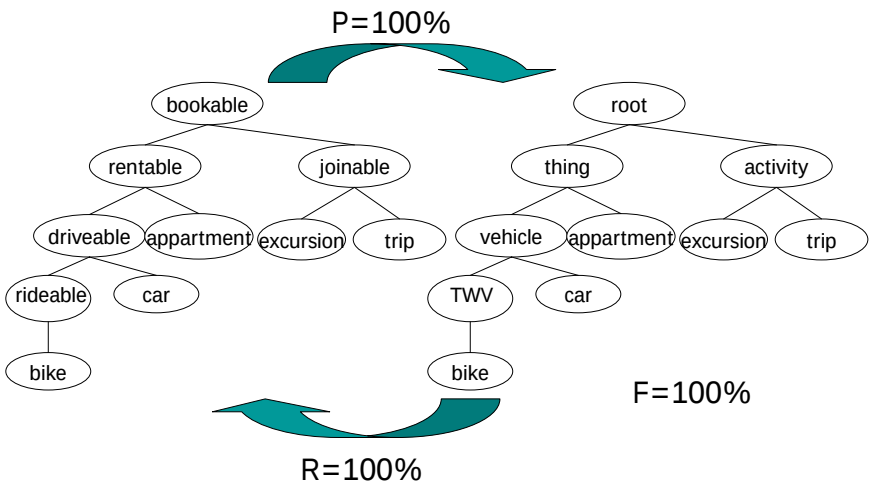


## Semantic Cotopy Revisited (Once More)

$$SC''(c, O_1, O_2) = \{c' \mid c' \in C_1 \cap C_2 \wedge (c' >_{O_1} c \vee c <_{O_1} c')\}$$

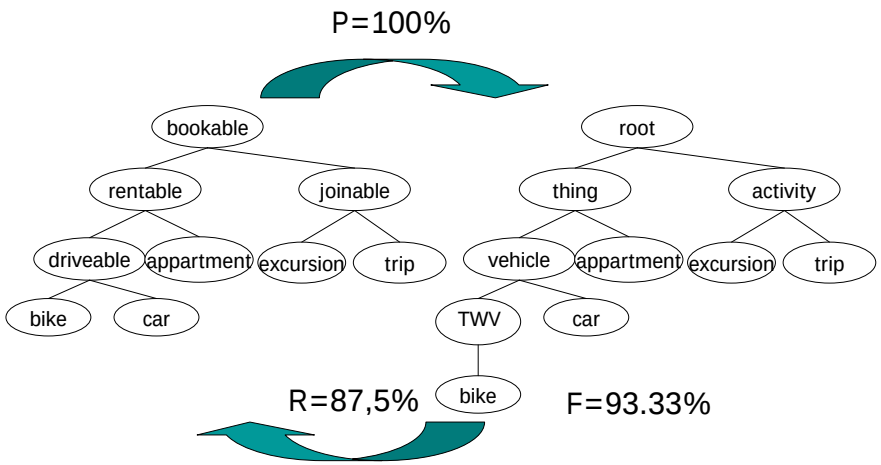
$$\overline{TO}(O_1, O_2) = \frac{1}{|C_1|} \sum_{c \in C_1, \notin C_2} TO(c, O_1, O_2)$$

# Example for Precision/Recall



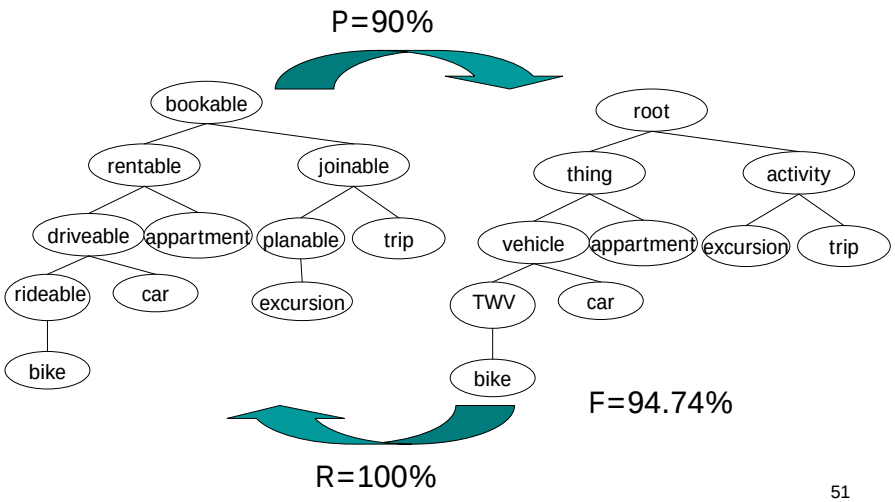



# Example for Precision/Recall



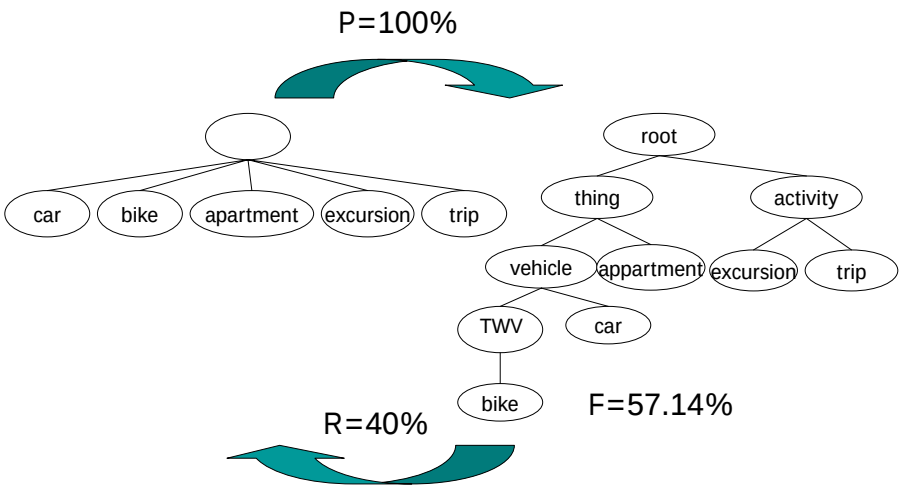



# Example for Precision/Recall






# Another Example



# Evaluation Methodology

$$\overline{TO}(O_1,O_2) = \frac{1}{|C_1|}\sum_{c\in C_1} TO(c,O_1,O_2)$$

$$TO(c,O_1,O_2) = \begin{cases} TO'(c,O_1,O_2)\ if\ c \in C_2 \\ TO''(c,O_1,O_2)\ if\ c \notin C_2 \end{cases}$$

$$TO'(c,O_1,O_2) := \frac{|SC(c,O_1,O_2) \cap SC(c,O_2,O_1)|}{|SC(c,O_1,O_2) \cup SC(c,O_2,O_1)|}$$

$$TO''(c,O_1,O_2) := \max_{c' \notin C_2} \frac{|SC(c,O_1,O_2) \cap SC(c',O_2,O_1)|}{|SC(c,O_1,O_2) \cup SC(c',O_2,O_1)|}$$

$$P(O_1,O_2) = \overline{TO}(O_1,O_2)$$

$$R(O_1,O_2) = \overline{TO}(O_2,O_1)$$

$$F(O_1,O_2) = \frac{2 \cdot P(O_1,O_2) \cdot R(O_1,O_2)}{P(O_1,O_2) + R(O_1,O_2)}$$



# Lexical Recall and F`

$$LR(O_1,O_2) = \frac{|C_{O_1} \cap C_{O_2}|}{|C_{O_2}|}$$

$$F'(O_1,O_2) = \frac{2 * F(O_1,O_2) * LR(O_1,O_2)}{(F(O_1,O_2) + LR(O_1,O_2))}$$



# Evaluation of Ontology Learning

- The aposteriori Approach:
  - ask domain expert for a per concept evaluation of the learned ontology
  - Count three categories of concepts:
    - *Correct* : both in learned and the gold ontology
    - *New* : only in learned ontology, but relevant and should be in gold standard as well
    - *Spurious*: useless
  - Compute precision = (*correct + new*) / (*correct + new + spurious*)
- As the result:
  The a priori evaluations are aweful – BUT
  A posteriori evaluations by domain experts still show very good results, very helpful for domain expert!

Sabou M., Wroe C., Goble C. and Mishne G.,Learning Domain Ontologies for Web Service Descriptions: an Experiment in Bioinformatics, In Proceeedings of the 14th International World Wide Web Conference (WWW2005), Chiba, Japan, 10-14 May, 2005.



# Starting Point in OL from text

- Context-based approaches:
  - Distributional Hypothesis [Harris 85]:
    „Words are (semantically) similar to the extent to which they appear in similar (syntactic) contexts“
  - leads to creation of groups
- Looking for explicit information:
  - Texts
  - WWW
  - Thesauri

## Looking for explicit information

There are two sources:

- Looking for patterns in texts:
  - ‚is-a' patterns [Hearst 92,98],[Poesio et al. 02], [Ahmid et al. 03]
  - ‚part-of' patterns [Charniak et al. 99]
  - ‚causation' patterns [Girju 02/03]
- Using the Web:
  - [Etzioni et al. 04]
  - [Cimiano et al. 04]



## Pattern based approaches (Hearst Patterns)

- Match patterns in corpus:
- NP0 such as NP1 ... NPn-1 (and|or) NPn
- such NP0 as NP1 ... NPn-1 (and|or) NPn
- NP1 ... NPn (and|or) other NP0
- NP0, (including,especially) NP1 ... NPn-1 (and|or) NPn

$$\text{for all } NP_i \; 1 \le i \le n \;\; \text{isa}_{Hearst}(\text{head}(NP_i), \text{head}(NP_0))$$

$$\text{isa}_{Hearst}(t_1, t_2) = \frac{\#\text{HearstPatterns}(t_1, t_2)}{\#\text{HearstPatterns}(t_1, *)}$$

- isaHearst(conference,event)=0.44
- isaHearst(conference,body)=0.22
- isaHearst(conference,meeting)=0.11
- isaHearst(conference,course)=0.11
- isaHearst(conference,activity)=0.11



## WWW Patterns

Generate patterns:

- <$t_1$>s such as <$t_2$>
- such <$t_1$>s as <$t_2$>
- <$t_1$>s, especially <$t_2$>
- <$t_1$>s, including <$t_2$>
- <$t_2$> and other <$t_2$>s
- <$t_2$> or other <$t_2$>s

and Query the Web using the GoogleAPI:

$$\text{isa}_{WWW}(t_1, t_2) = \frac{\#\text{Patterns}(t_1, t_2)}{\#\text{Patterns}(t_1, *)}$$



## The Vector-Space Model

- **Idea:** collect context information based on the distributional hypothesis and represent it as a vector:

| | die_from | suffer_from | enjoy | eat |
|---|---|---|---|---|
| disease | X | X | | |
| cirrhosis | X | X | | |

- compute similarity among vectors wrt. to some measure

## Clustering Concept Hierarchies from Text

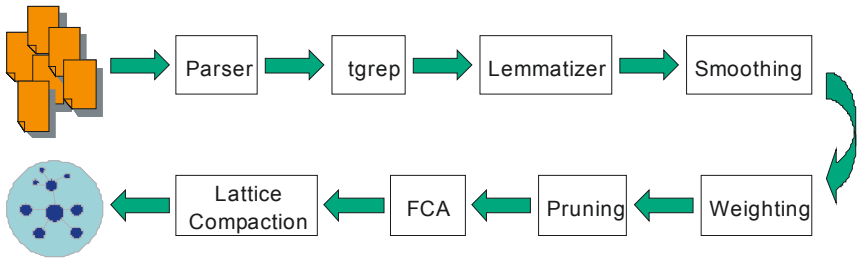
- Observation: ontology engineers need information about the effectiveness, efficiency and trade-offs of different approaches
- Similarity-based
  - agglomerative/bottom-up
  - divisive/top-down: Bi-Section-KMeans
- Set-theoretical
  - set operations (inclusion)
  - FCA, based on Galois lattices

**[Cimiano et al. 03, 04]**



## Context Extraction

- extract syntactic dependencies from text
  ⇒ verb/object, verb/subject, verb/PP relations
  ⇒ car: drive_obj, crash_subj, sit_in, ...
- LoPar, a trainable statistical left-corner parser:

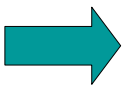





## Example

- People book hotels. The man drove the bike along the beach.

book_subj(people)
book_obj(hotels)
drove_subj(man)
drove_obj(bike)
drove_along(beach)




book_subj(people)
book_obj(hotel)
drive_subj(man)
drive_obj(bike)
drive_along(beach)



## Weighting (threshold t)

- Conditional: $P(n \mid v_{\arg})$
- Hindle: $P(n \mid v_{\arg}) \cdot \log\left(\frac{P(n \mid v_{\arg})}{P(n)}\right)$
- Resnik: $S_R(v_{\arg}) \cdot P(n \mid v_{\arg}) \cdot \log\left(\frac{P(n \mid v_{\arg})}{P(n)}\right)$

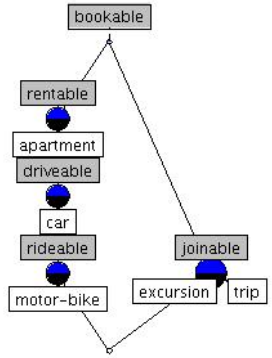
$$S_R(v_{\arg}) = \sum_{n'} P(n' \mid v_{\arg}) \cdot \log\left(\frac{P(n' \mid v_{\arg})}{P(n')}\right)$$

# Tourism Formal Context

| | bookable | rentable | driveable | rideable | joinable |
|---|---|---|---|---|---|
| appartment | X | X | | | |
| car | X | X | X | | |
| motor-bike | X | X | X | X | |
| excursion | X | | | | X |
| trip | X | | | | X |

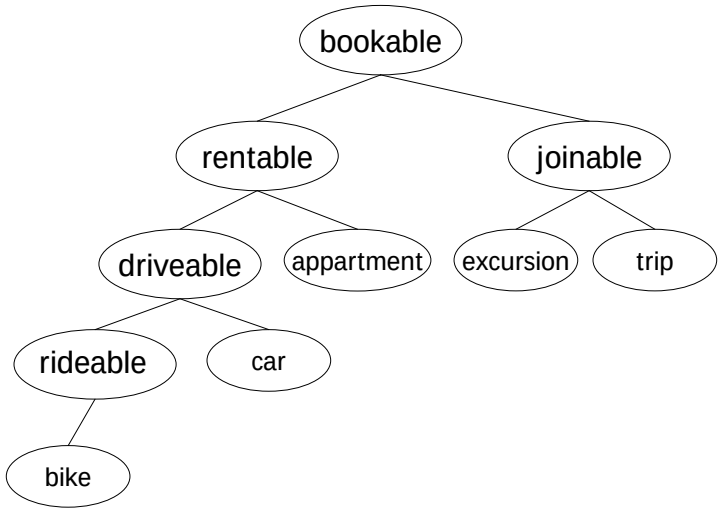


# Tourism Lattice



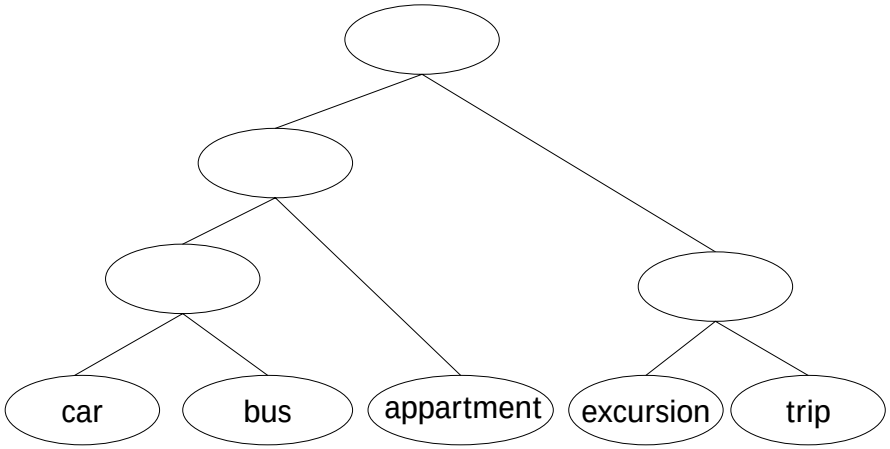



# Concept Hierarchy






# Agglomerative/Bottom-Up Clustering

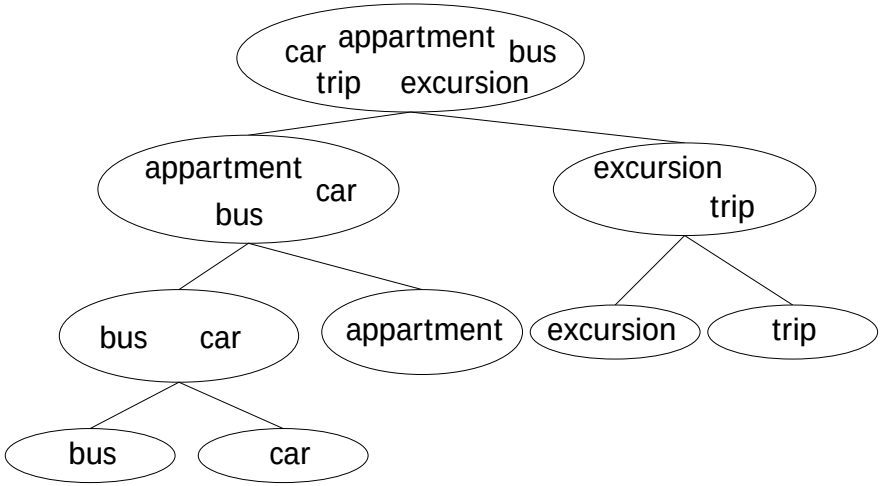


# Linkage Strategies

- Complete-Linkage:
  - consider the two most dissimilar elements of each of the clusters => $O(n^2 \log(n))$
- Average-Linkage:
  - consider the average similarity of the elements in the clusters => $O(n^2 \log(n))$
- Single-Linkage:
  - consider the two most similar elements of each of the clusters => $O(n^2)$



# Bi-Section-KMeans



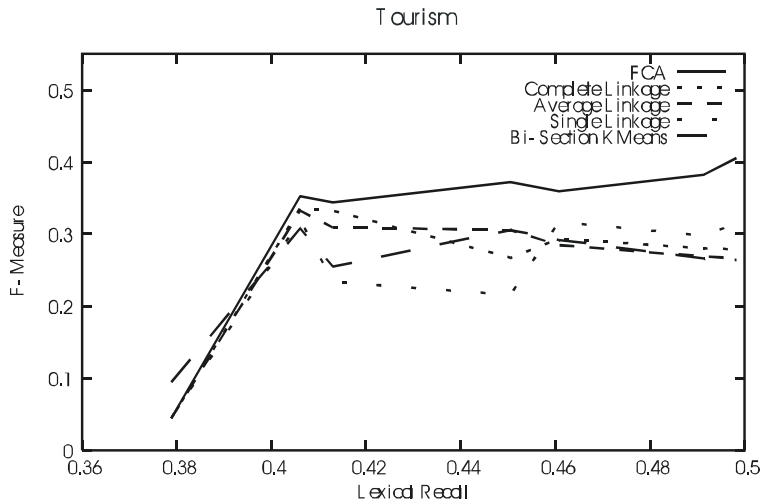



# Data Sets

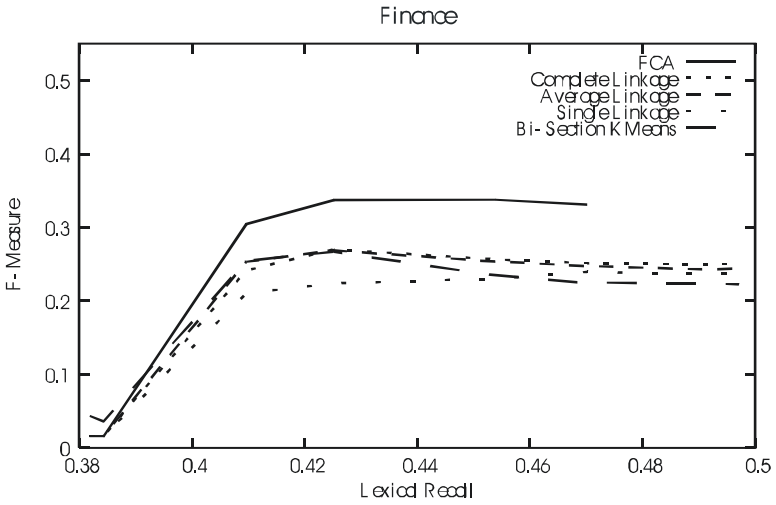
- Tourism (118 Mio. tokens):
  - http://www.all-in-all.de/english
  - http://www.lonelyplanet.com
  - British National Corpus (BNC)
  - handcrafted tourism ontology (289 concepts)
- Finance (185 Mio. tokens):
  - Reuters news from 1987
  - GETESS finance ontology (1178 concepts)



# Results Tourism Domain

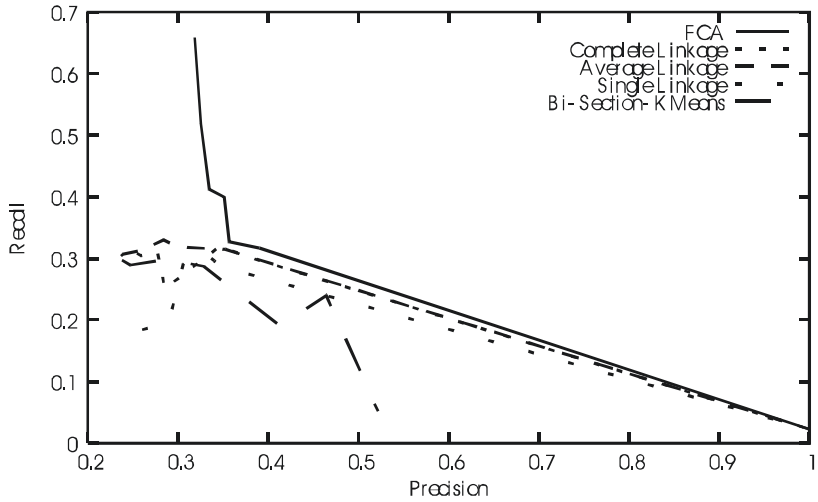


# Results in Finance Domain



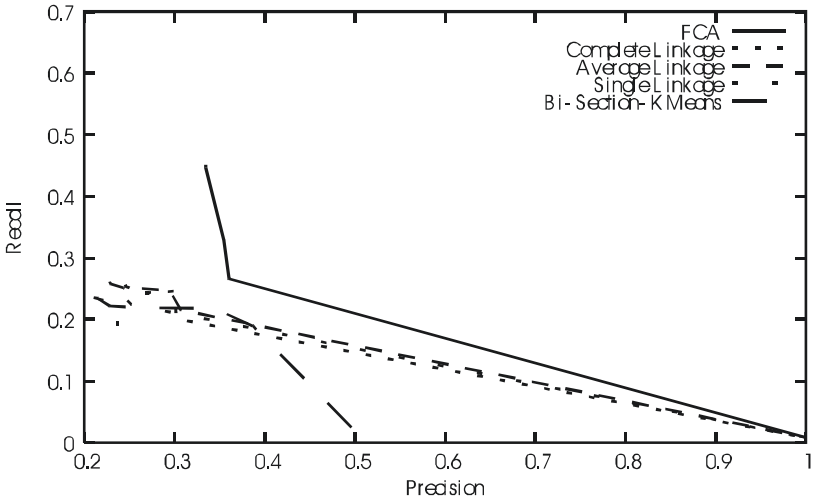



# Results Tourism Domain






# Results in Finance Domain






# Summary

| | Effectiveness | Efficiency | Traceability |
|---|---|---|---|
| FCA | 43.81/41.02% | $O(2^n)$ | Good |
| Agglomerative Clustering | 36.78/33.35%<br>36.55/32.92%<br>38.57/32.15% | $O(n^2 \log(n))$<br>$O(n^2 \log(n))$<br>$O(n^2)$ | Fair |
| Divisive Clustering | 36.42/32.77% | $O(n^2)$ | Weak-Fair |

# Other Clustering Approaches

- Bottom-Up/Agglomerative
  - (ASIUM System) Faure and Nedellec 1998
  - Caraballo 1999
  - (Mo'K Workbench) Bisson et al. 2000
- Other:
  - Hindle 1990
  - Pereira et al. 1993
  - Hovy et al. 2000



# Ontology Learning References


- Reinberger, M.-L., & Spyns, P. (2005). Unsupervised text mining for the learning of dogma-inspired ontologies. In Buitelaar, P., Cimiano, P., & Magnini, B. (Eds.), Ontology Learning from Text: Methods, Evaluation and Applications.
- Philipp Cimiano, Andreas Hotho, Steffen Staab: Comparing Conceptual, Divise and Agglomerative Clustering for Learning Taxonomies from Text. ECAI 2004: 435-439
- P. Cimiano, A. Pivk, L. Schmidt-Thieme and S. Staab, Learning Taxonomic Relations from Heterogenous Evidence. In Buitelaar, P., Cimiano, P., & Magnini, B. (Eds.), Ontology Learning from Text: Methods, Evaluation and Applications.
- Sabou M., Wroe C., Goble C. and Mishne G.,Learning Domain Ontologies for Web Service Descriptions: an Experiment in Bioinformatics, In Proceeedings of the 14th International World Wide Web Conference (WWW2005), Chiba, Japan, 10-14 May, 2005.
- Alexander Maedche, Ontology Learning for the Semantic Web, PhD Thesis, Kluwer, 2001.
- Alexander Maedche, Steffen Staab: Ontology Learning for the Semantic Web. IEEE Intelligent Systems 16(2): 72-79 (2001)
- Alexander Maedche, Steffen Staab: Ontology Learning. Handbook on Ontologies 2004: 173-190
- M. Ciaramita, A. Gangemi, E. Ratsch, J. Saric, I. Rojas. Unsupervised Learning of semantic relations between concepts of a molecular biology ontology. IJCAI, 659ff.
- A. Schutz, P. Buitelaar. RelExt: A Tool for Relation Extraction from Text in Ontology Extension. ISWC 2005.
- Faure, D., & N´edellec, C. (1998). A corpus-based conceptual clustering method for verb frames and ontology. In Velardi, P. (Ed.), Proceedings of the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications, pp. 5–12.
- Michele Missikoff, Paola Velardi, Paolo Fabriani: Text Mining Techniques to Automatically Enrich a Domain Ontology. Applied Intelligence 18(3): 323-340 (2003).
- Gilles Bisson, Claire Nedellec, Dolores Cañamero: Designing Clustering Methods for Ontology Building - The Mo'K Workbench. ECAI Workshop on Ontology Learning 2000




# Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications



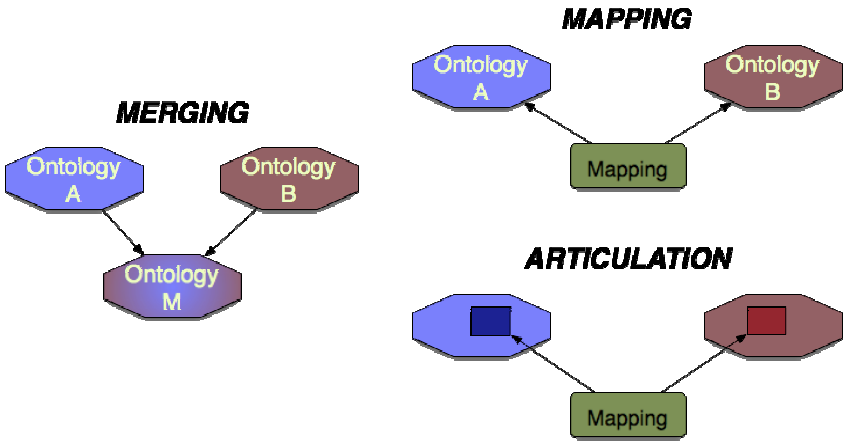
# Lots of Overlapping Ontologies on the Semantic Web


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


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



## Creating Correspondences Between Ontologies







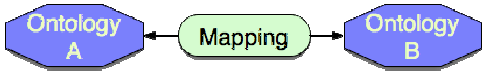

## Ontology-level Mismatches

- 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-to-Ontology Mappings: Sources of information



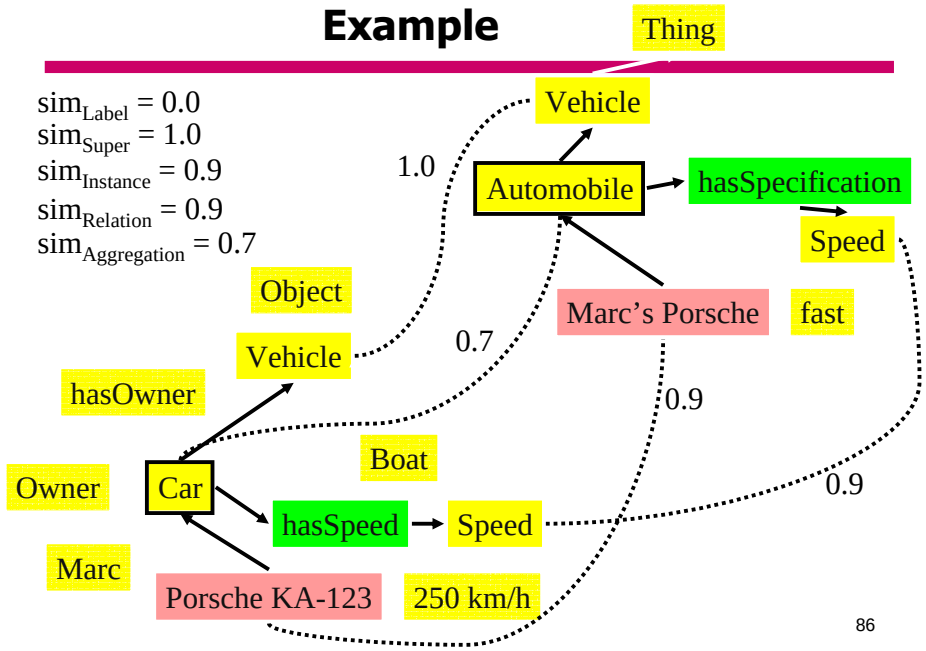

- Lexical information: edit distance, …
- Ontology structure: subclassOf, instanceOf,…
- User input: “anchor points”
- External resources: WordNet,…
- Prior matches

## Mapping Methods

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



## Example



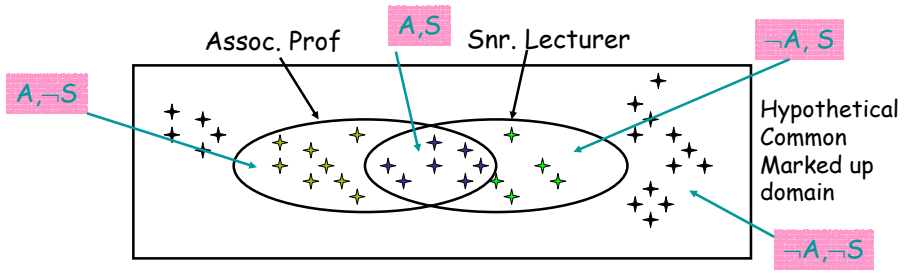



## Mapping Methods

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



## GLUE: Defining Similarity




$$Sim(\text{Assoc. Prof., Snr. Lect.}) = \frac{P(\mathbf{A} \cap \mathbf{S})}{P(\mathbf{A} \cup \mathbf{S})} = \frac{P(A,S)}{P(A,\neg S) + P(A,S) + P(\neg A,S)}$$

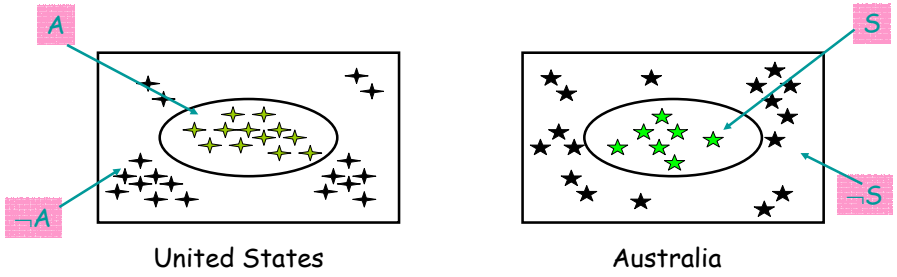
[Jaccard, 1908]

Joint Probability Distribution: P(A,S),P(¬A,S),P(A,¬S),P(¬A,¬S)

**Multiple Similarity measures in terms of the JPD**

## GLUE: No common data instances

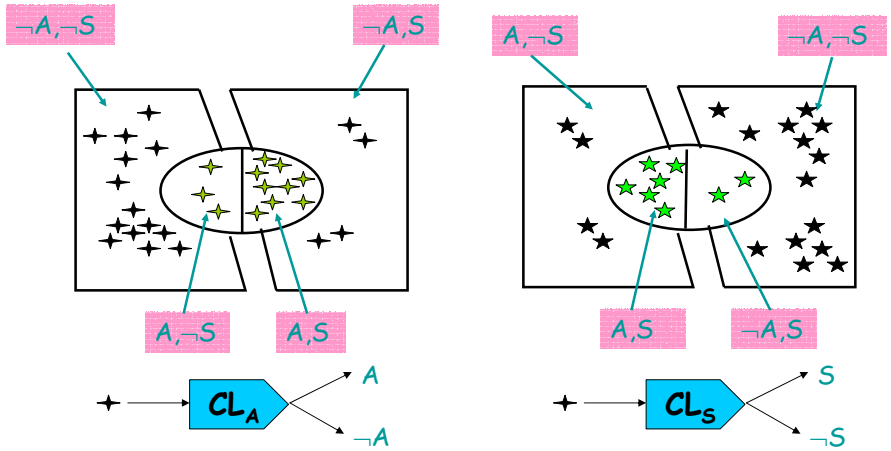
In practice, not easy to find data tagged with both ontologies !




Solution: Use Machine Learning



## Machine Learning for computing similarities



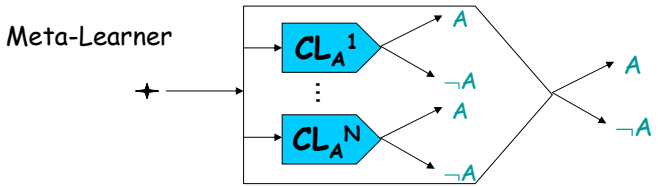

JPD estimated by counting the sizes of the partitions



## GLUE: Improve Predictive Accuracy – Use Multi-Strategy Learning

Single Classifier cannot exploit all available information

Combine the prediction of multiple classifiers




Content Learner

Frequencies on different words in the text in the data instances

Name Learner

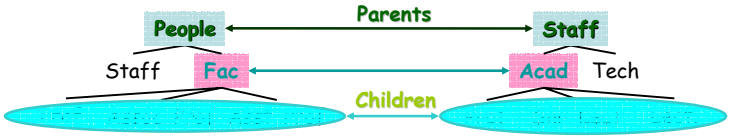
Words used in the names of concepts in the taxonomy

Others …



## GLUE Next Step: Exploit Constraints

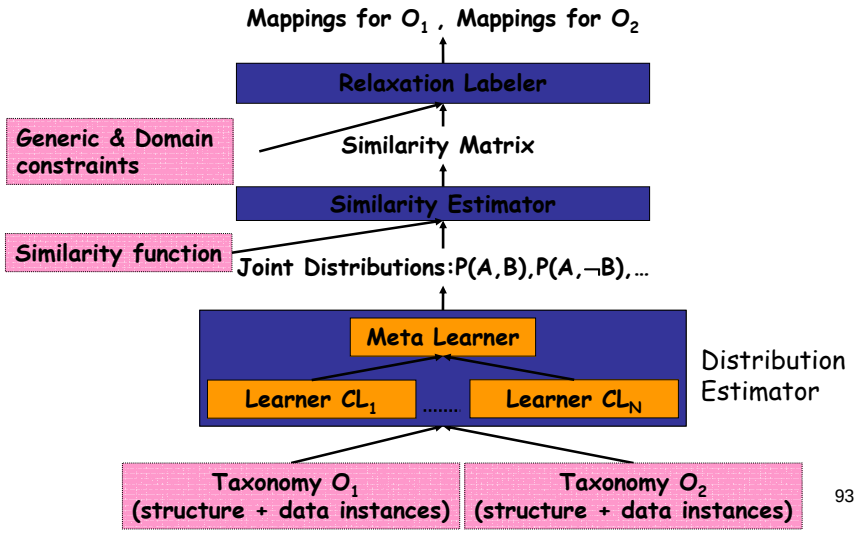
- Constraints due to the taxonomy structure




- Domain specific constraints
  - *Department-Chair* can only map to a unique concept
- Numerous constraints of different types

Extended Relaxation Labeling to ontology matching

# Putting it all together GLUE System






# APFEL: Similarity Features

| | *Feature* | *Similarity Measure* |
|---|---|---|
| Concepts | label | String Similarity |
| | subclassOf | Set Similarity |
| | instances | Set Similarity |
| | ... | |
| Relations | | |
| Instances | | |

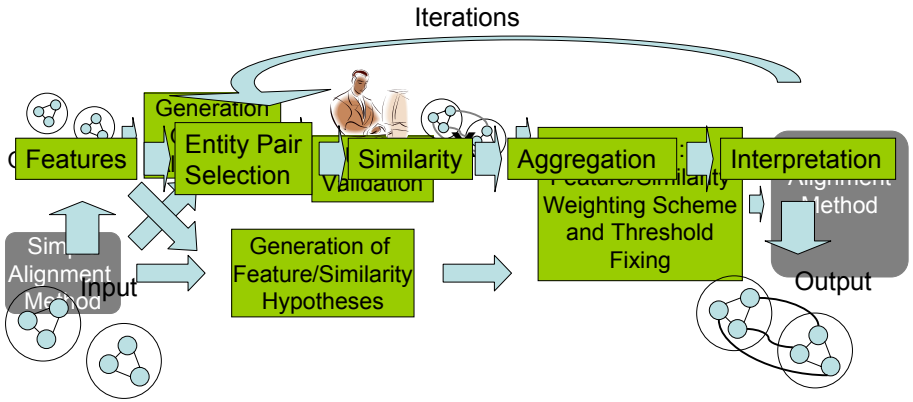
Aggregation - Example:

$$sim(e, f) = \sum_k w_k sim_k(e, f)$$

Interpretation: $map(e_{1j}) = e_{2j} \leftarrow sim(e_{1j}, e_{2j}) > t$

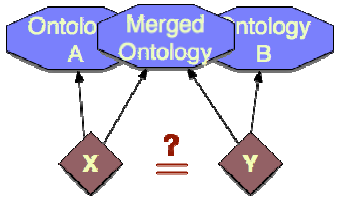


# APFEL: Optimize Integration





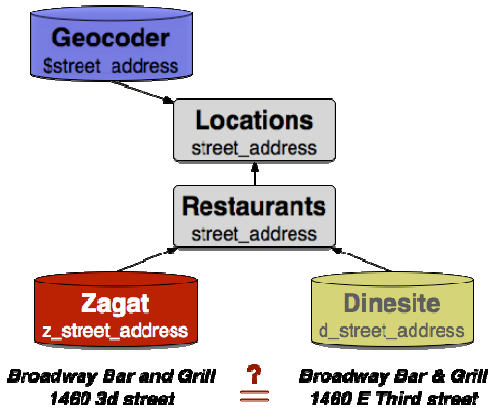

# Duplicate Recognition




- Do two objects refer to the same entity?
  - We know objects have the same type (their types are mapped/merged)
- Examples
  - Duplicate removal after merging knowledge bases
  - Citation matching

## Using External Sources for Duplicate Recognition

- Appolo (USC/ISI)
  - Combines information-integration mediator (Prometheus) with a record-linkage system (Active Atlas)
  - Uses a domain model of sources and information that they provide

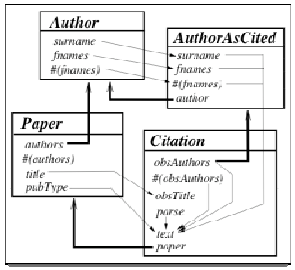







## Duplicate Recognition: Citation Matching

- Pasula, Marthi, et.al. (UC Berkeley)
  - Performs citation matching based on probability models for
    - author names
    - titles
    - title corruption, etc.
  - Extends standard domain model to incorporate probabilities
  - Learns probability models from large data sets








## References


- User Input driven - Prompt, Chimaera, ONION
- Chimaera (Stanford KSL; D. McGuinness et al)

- AnchorPrompt (Stanford SMI; Noy, Musen et al)
- Similarity Flooding (Melnik, Garcia-Molina, Rahm)

- IF-Map (Kalfoglou, Schorlemmer)

- Using metrics to compare OWL concepts (Euzenat and Volchev)
- QOM (Ehrig and Staab)

- Corpus of Matches (O.Etzioni, A. Halevy, et.al.)
- APFEL (Ehrig, Staab, Sure)

- SAT Reasoning - S-Match (U. Trento; Serafini et al)

- Mapping Composition: Semantic gossiping (Aberer et al), Piazza (Halevy et al), Prasenjit Mitra

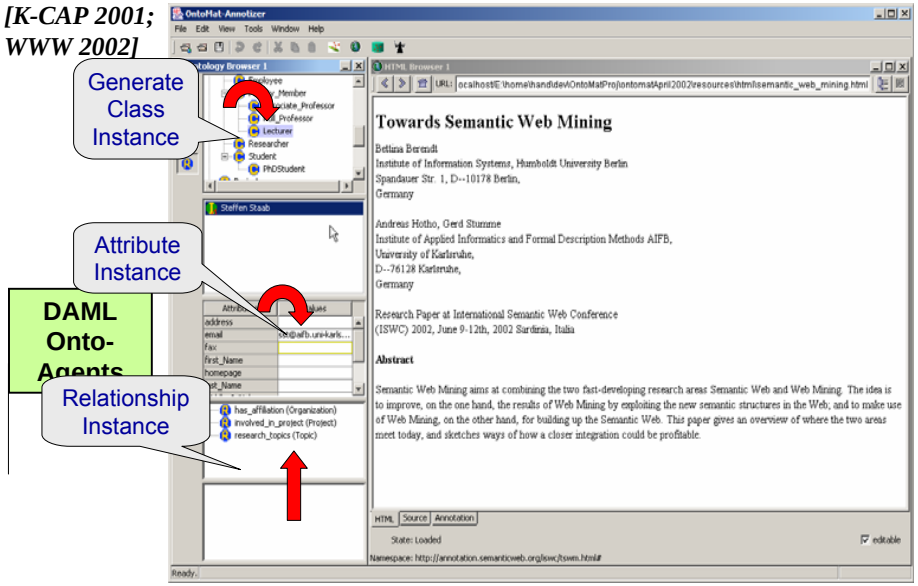



## Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications

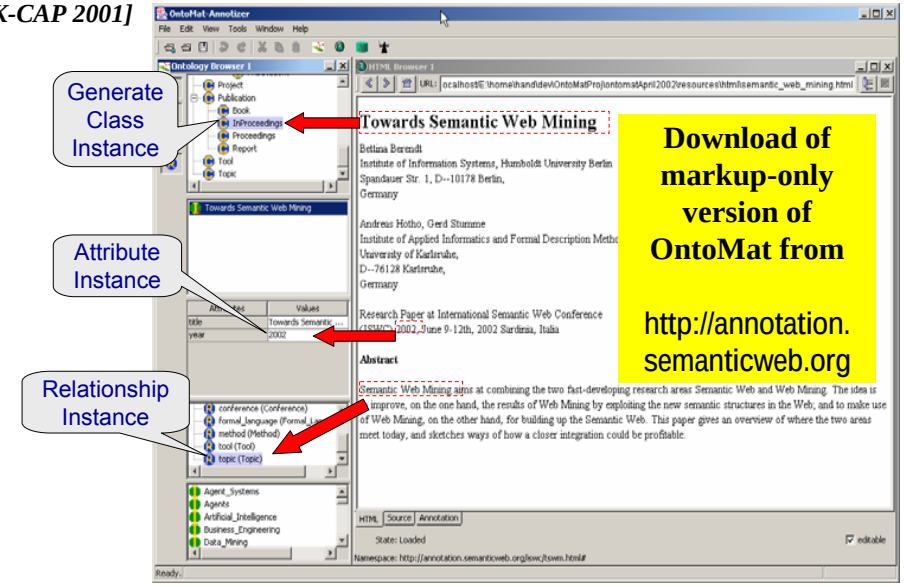
# CREAM – Creating Metadata

*[K-CAP 2001; WWW 2002]*



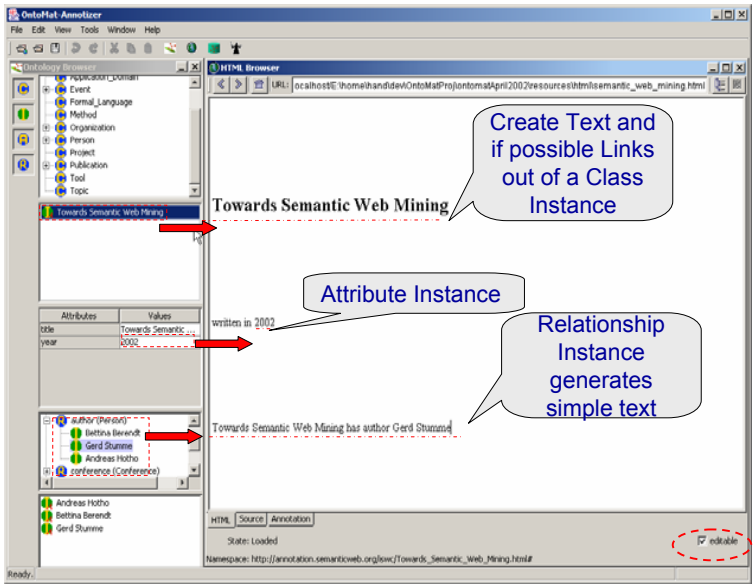

# Annotation by Markup

*[K-CAP 2001]*

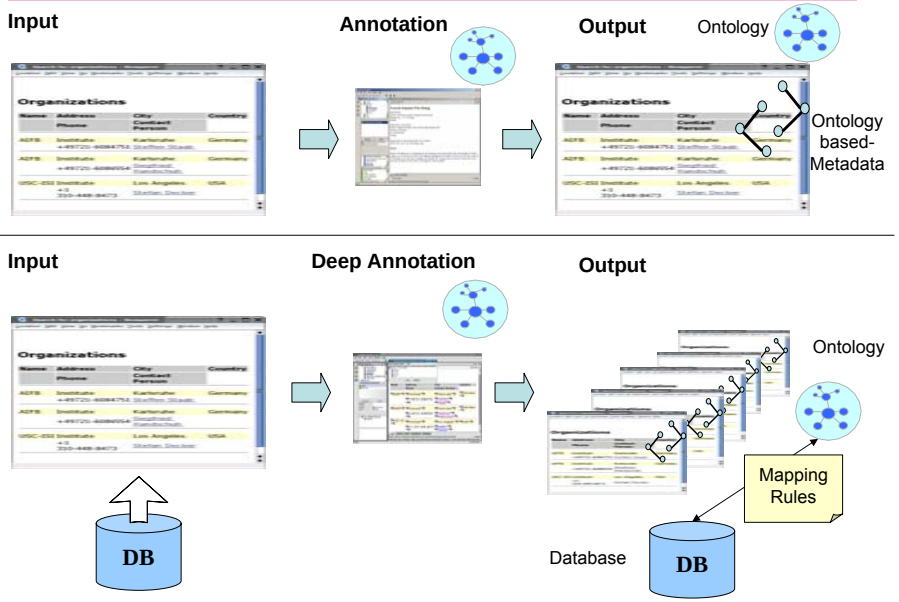



# Annotation by Authoring

*[WWW 2002]*




# *[WWW 2003]* Annotation vs. Deep Annotation

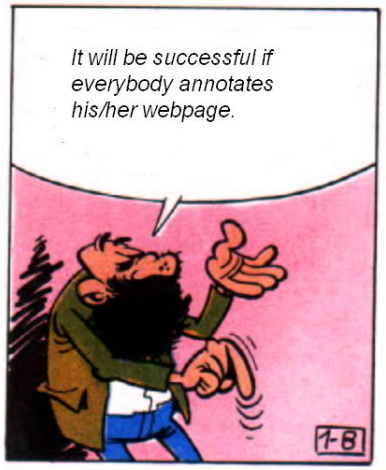


## The annotation problem in 4 cartoons



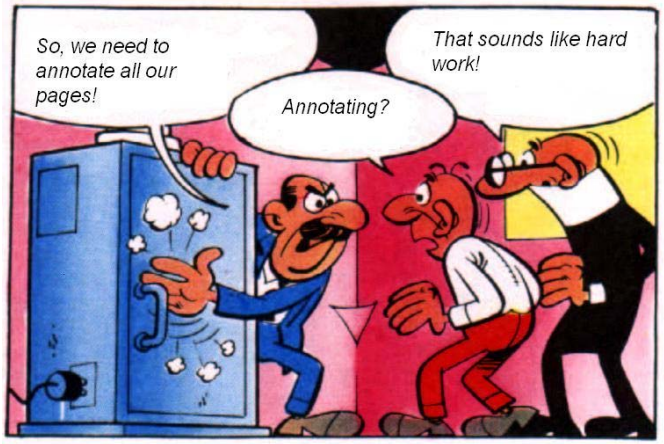



## The annotation problem from a scientific point of view






## The annotation problem in practice






## The vicious cycle

# Current State-of-the-art

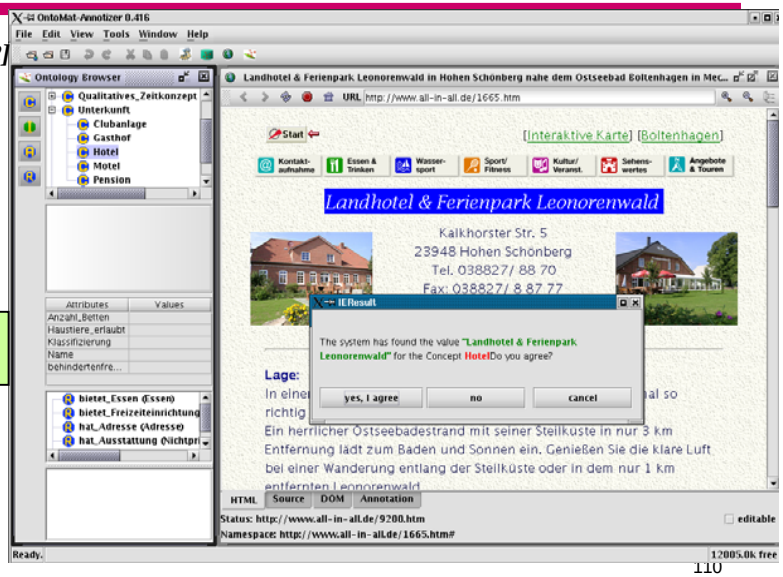
- ML-based IE (e.g.Amilcare@{OntoMat,MnM})
  - start with hand-annotated training corpus
  - rule induction
- Standard IE (MUC)
  - handcrafted rules
  - Wrappers
- Large-scale IE [SemTag&Seeker@WWW'03]
  - Large scale system
  - disambiguation with TAP
- (C-)Pankow (Cimiano et.al. WWW'04, WWW'05)
- KnowItAll (Etzioni et al. WWW'04)



# Semi-automatic Annotation

*[EKAW 2002]*

**EU IST Dot-Kom**

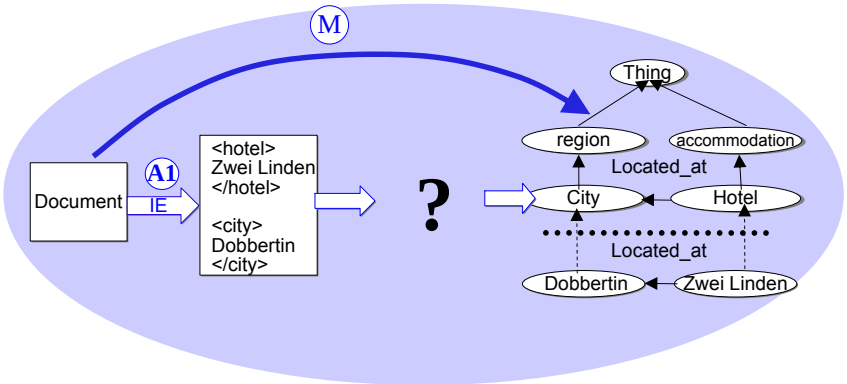





# Comparison of CREAM and S-CREAM

Core processes: Input, Output

- (M) Manual Annotation (OntoMat) → Relational Metadata
- (A1) Information Extraction (Amilcare) → XML annotated Dokument





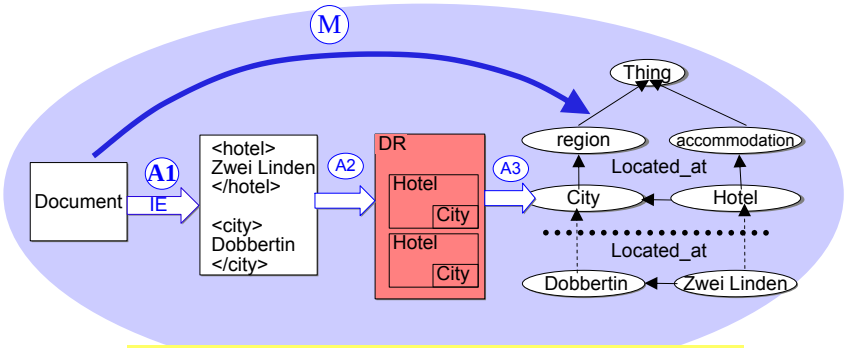

# Different Results

| | |
|---|---|
| <hotel> Zwei Linden </hotel> | Zwei Linden **InstOf** Hotel |
| | Zwei Linden **Locatet_At** Dobbertin |
| <city>Dobbertin</city> | Dobbertin **InstOf** City |
| | Zwei Linden **Has_Room** single_room_1 |
| <singleroom>Single room</singleroom> | single_room1 **InstOf** Single_Room |
| | single_room1 **Has_Rate** rate1 |
| | rate1 **InstOf** Rate |
| <price>25,66</price> | rate1 **Price** 25,66 |
| <currency>EUR</currency> | rate1 **Currency** EUR |
| | Zwei Linden **Has_Room** double_room1 |
| <doubleroom>Double room</doubleroom> | double_room1 **InstOf** Double_Room |
| | double_room1 **Has_Rate** rate2 |
| | rate2 **InstOf** Rate |
| <price>43,66</price> | rate2 **Price** 43,46 |
| <currency>EUR</currency> | rate2 **Currency** EUR |
| Amilcare (IE-Tool) | OntoMat-Annotizer |

# Comparison of CREAM and S-CREAM

Core processes: Input, Output

- (M) Manual Annotation (OntoMat) → Relational Metadata
- (A1) Information extraction (Amilcare) → XML annotated Document




*Currently: Simple Centering-Modell*

*Future: Learn Coherency Rules*



# IE and Wrapper Learning

- Boosted wrapper induction
- Exploiting linguistic constraints
- Hidden Markov models
- Data mining and IE
- Bootstrapping
- First-order learning



# Wrapper

No tutorial about IE and Wrapper learning but...

- IE often focuses on small number of classes
- Is not easily adaptable to new domains
- Needs a lot of trainings examples

Needed

- It would be great if IE would scale to a large number of classes (concepts) on a large amount of unlabeled data



# SemTag

- The goal is to add semantic tags to the existing HTML body of the web.
- SemTag uses TAP, where TAP is a public broad, shallow knowledgebase.
- TAP Contains lexical and taxonomical information about popular objects like music, movies, sports, etc.

Example:

"The Chicago Bulls announced that Michael Jordan will..."

Will be:

The <resource ref = http://tap.stanford.edu/Basketball Team_Bulls>Chicago Bulls</resource> announced yesterday that <resource ref = "http://tap.stanford.edu/ AthleteJordan_Michael"> Michael Jordan</resource> will..."



**Dill et al, SemTag and Seeker. WWW'03**

## SemTag

- Lookup of all instances from the ontology (TAP) – 65K instances
- Disambiguate the occurrences as:
  - One of those in the taxonomy
  - Not present in the taxonomy
- Placing labels in the taxonomy is hard
- Use bag-of-words approach for disambiguation
- 3 people evaluated 200 labels in context – agreed on only 68.5% - metonymy
- Applied on 264 million pages
- Produced 550 million labels and 434 spots
- Accuracy 82%



**Dill et al, SemTag and Seeker. WWW'03**

## The Self-Annotating Web

- There is a huge amount of non-formalized knowledge in the Web
- Use statistics to interpret this non-formalized knowledge and propose formal annotations:

**semantics ≈ syntax + statistics?**

- Annotation by maximal statistical evidence



## **PANKOW:** P**attern-based** AN**notation through** K**nowledge** O**n the** W**eb**

- HEARST1: <CONCEPT>s such as <INSTANCE>
- HEARST2: such <CONCEPT>s as <INSTANCE>
- HEARST3: <CONCEPT>s, (especially/including) <INSTANCE>
- HEARST4: <INSTANCE> (and/or) other <CONCEPT>s

- Examples:
  - countries such as Niger
  - such countries as Niger
  - countries, especially Niger
  - countries, including Niger
  - Niger and other countries
  - Niger or other countries

instanceOf(Niger,country)



## Patterns (Cont'd)

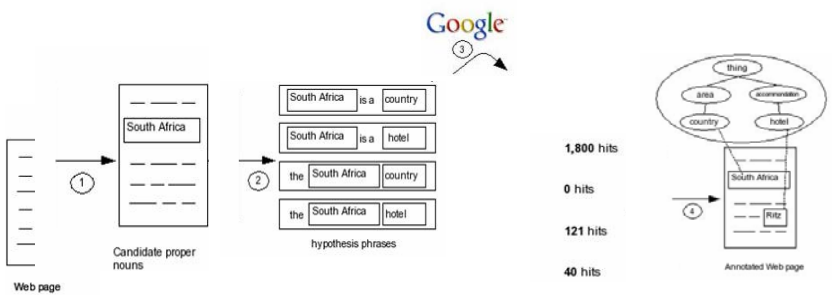
- DEFINITE1: the <INSTANCE> <CONCEPT>
- DEFINITE2: the <CONCEPT> <INSTANCE>

- APPOSITION: <INSTANCE>, a <CONCEPT>
- COPULA: <INSTANCE> is a <CONCEPT>

- Examples:
- the Niger country
- the country Niger
- Niger, a country in Africa
- Niger is a country in Africa

instanceOf(Niger,country)

# PANKOW Process






# Gimme‘ The Context: C-PANKOW

- **Contextualize the pattern-matching** by taking into account the similarity of the Google-abstract in which the pattern was matched and the one to be annotated
- **Download a fixed number n of Google-abstracts** matching so-called *clues* and analyze them linguistically, matching the patterns offline:
  - match more complex structures
  - more efficient as the number of Google-queries only depends on n
  - more offline processing, reducing network traffic



# Comparison

| System | # | Recall/ Accuracy | Learning Accuracy |
|---|---|---|---|
| [MUC-7] | 3 | >> 90% | n.a. |
| [Fleischman02] | 8 | 70.4% | n.a. |
| PANKOW | 59 | 24.9% | 58.91% |
| [Hahn98] –TH | 325 | 21% | 67% |
| [Hahn98]-CB | 325 | 26% | 73% |
| [Hahn98]-CB | 325 | 31% | 76% |
| C-PANKOW | 682 | 29.35% | 74.37% |
| [Alfonseca02] | 1200 | 17.39% (strict) | 44% |

LA based on least common superconcept lcs of two concepts (Hahn et.al. 98)



# Web-scale information extraction

KnowItAll Idea:

- Web is the largest knowledge base
- The goal is to find all instances corresponding to a given concept in the web and extract them

The System is:

- Domain-Independent
- Use Bootstrap technique
- Based on Linguistic Patterns

KnowItAll vs (C-)Pankow

- Pankow starts from a Web page and annotates a given term on the page using the Web
- KnowItAll starts from a concept and aims at finding all instances on the Web

## References Semantic Annotation


- S. Handschuh, S. Staab (eds.). Annotation for the Semantic Web. IOS Press, 2003
- P. Cimiano, S. Handschuh, S. Staab. Towards the Self-annotating Web. 13th International World Wide Web Conference, WWW 2004, New York, USA, May 17-22, 2004.
- Siegfried Handschuh, Creating Ontology-based Metadata by Annotation for the Semantic Web, PhD Thesis, 2005.
- O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D.S.Weld, and A. Yates. Web-scale information extraction in KnowItAll (preliminary results). In *Proceedings of the 13th World Wide Web Conference*, pages 100–109, 2004.
- S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, T. Kanungo, S. Rajagopalan, A. Tomkins, J.A. Tomlin, and J.Y. Zien. Semtag and seeker: bootstrapping the semantic web via automated semantic annotation. In *Proceedings of the 12th International World Wide Web Conference*, pages 178–186. ACM Press, 2003.
- S. Brin. Extracting patterns and relations from the World Wide Web. In Proceedings of the WebDB Workshop at EDBT '98, 1998.
- F. Ciravegna, A. Dingli, D. Guthrie, and Y. Wilks. Integrating Information to Bootstrap Information Extraction from Web Sites. In Proceedings of the IJCAI Workshop on Information Integration on the Web, pages 9–14, 2003.
- H. Cui, M.-Y. Kan, and T.-S. Chua. Unsupervised learning of soft patterns for generating definitions from online news. In Proceedings of the 13th World Wide Web Conference, pages 90–99, 2004.
- U. Hahn and K. Schnattinger. Towards text knowledge engineering. In AAAI'98/IAAI'98 Proceedings of the 15th National Conference on Artificial Intelligence and the 10th Conference on Innovative Applications of Artificial Intelligence, 1998




## Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications



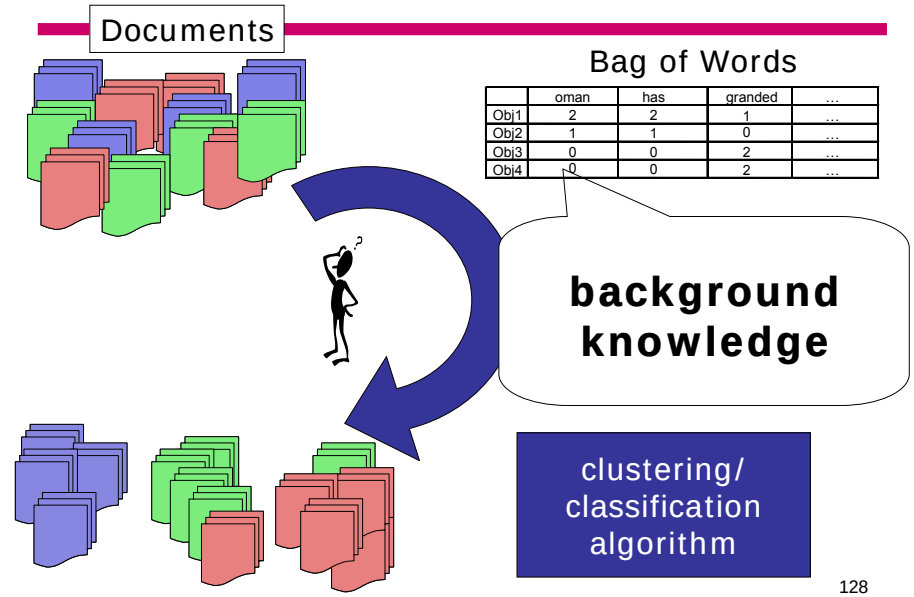
## Using Ontologies

Ontologies as:

- background knowledge for text clustering and classification
- basis for recommender systems
- background knowledge in ILP
- knowledge for models in Statistical Relational Learning



## Text Clustering & Classification Approaches



# Text Clustering & Classification Approaches

Documents

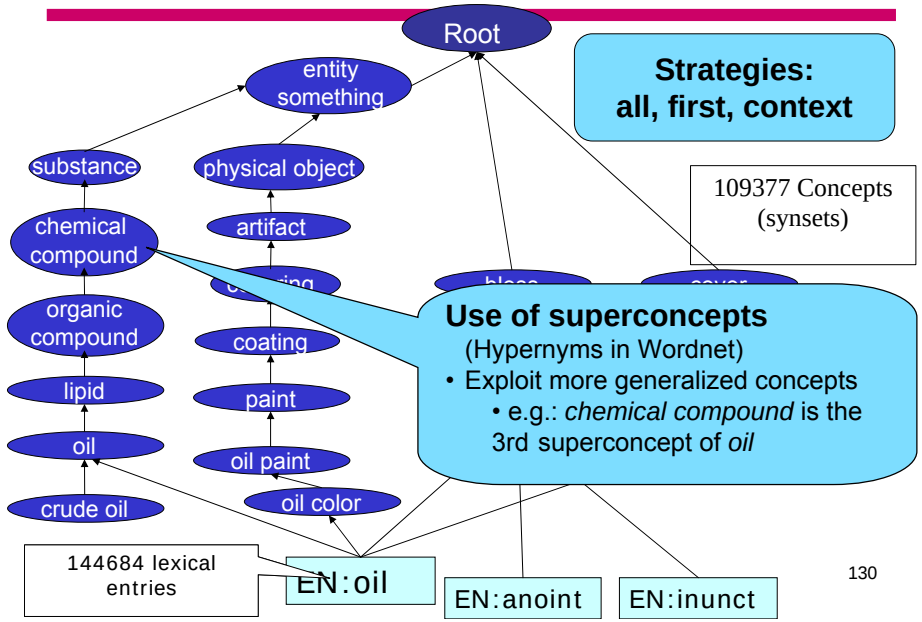
Dok 17892 crude
==========
Oman has granted term crude oil customers retroactive discounts from official prices of 30 to 38 cents per barrel on liftings made during February, March and April, the weekly newsletter Middle East Economic Survey (MEES) said. MEES said the price adjustments, arrived at through negotiations between the Omani oil ministry and companies concerned, are designed to compensate for the difference between market-related prices and the official 17.63 dlrs per barrel adopted OPEC Oman since February.
REUTER

**Bag of Words**

| | |
|---|---|
| Oman | 2 |
| has | 1 |
| granted | 1 |
| term | 1 |
| crude | 1 |
| oil | 2 |
| customers | 1 |
| retroactive | 1 |
| discounts | 1 |
| ... | ... |

Further preprocessing steps
-Stopwords
-Stemming

# WordNet as an example and ontology




130

# Ontology-based representation

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Oman | 1 | | Oman | 1 | | Oman | 1 |
| has | 1 | | granted | 1 | | granted | 1 |
| granted | 1 | | term | 1 | | term | 1 |
| term | 1 | | (C) term | 1 | | (C) term | 1 |
| crude | 1 | | crude | 1 | | crude | 1 |
| oil | 1 | | (C) crude | 1 | | (C) crude | 1 |
| customers | 1 | | oil | 1 | | oil | 1 |
| retroactive | 1 | | (C) oil | 1 | | (C) oil | 1 |
| discounts | 1 | | customer | 1 | | (C) lipid | 1 |
| ... | ... | | (C) customer | 1 | | (C) compound | 1 |
| | | | ... | ... | | ... | ... |

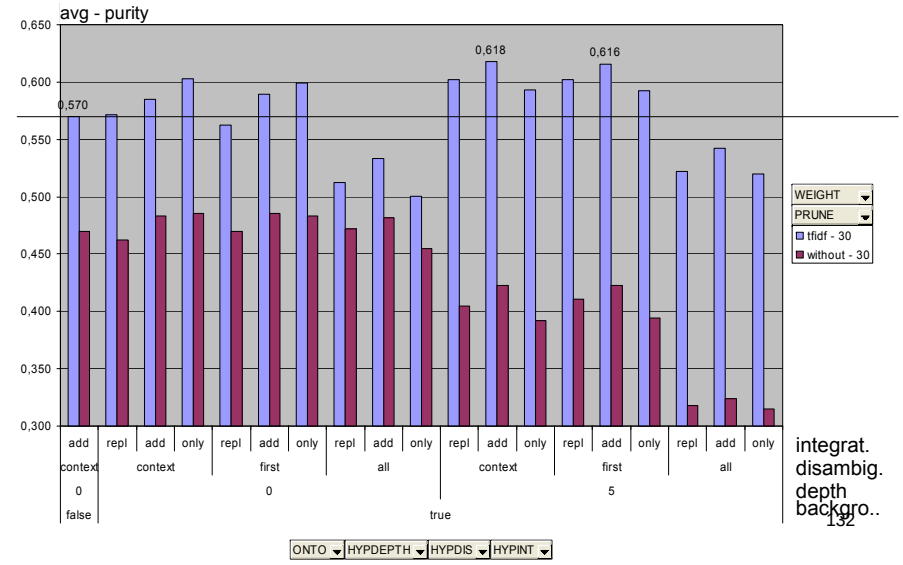
1 2 3

131

strategy: add

# n of Text Clustering

**Evaluation parameter**
- min 15, max 100, 2619 documents of the reuters corpus
- cluster k = 60, with BiSec-KMeans




132

## Evaluation: OHSUMED Classification Results

Top 50 classes with WordNet and AdaBoost

| Feature Type | Error | macro-averaged Prec | Rec | $F_1$ | BEP |
|---|---|---|---|---|---|
| term | 00.53 | 52.60 | 35.74 | 42.56 | 45.68 |
| term & synset.first | **00.52** | 53.08 | 36.98 | 43.59 | 46.46 |
| term & synset.first.hyp5 | **00.52** | **53.82** | **38.66** | **45.00** | **48.01** |
| term & synset.context | **00.52** | 52.83 | 37.09 | 43.58 | **46.88** |
| term & synset.context.hyp5 | **00.51** | **54.55** | **39.06** | **45.53** | **48.10** |
| term & synset.all | **00.52** | 52.89 | 37.09 | 43.60 | 46.82 |
| term & synset.all.hyp5 | **00.52** | **53.33** | **38.24** | **44.42** | 46.73 |

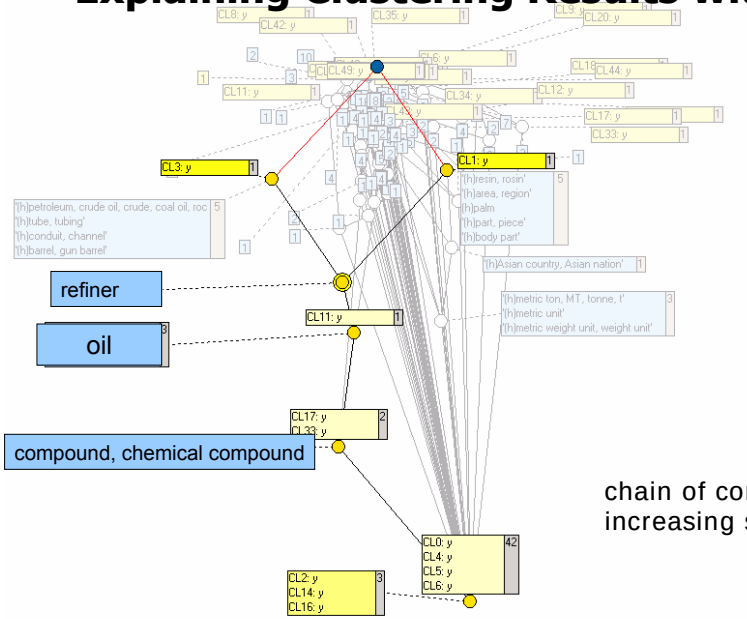
| Feature Type | Error | micro-averaged Prec | Rec | $F_1$ | BEP |
|---|---|---|---|---|---|
| term | 00.53 | 55.77 | 36.25 | 43.94 | 46.17 |
| term & synset.first | **00.52** | 56.07 | 37.30 | 44.80 | 47.01 |
| term & synset.first.hyp5 | **00.52** | **56.84** | **38.76** | **46.09** | **48.31** |
| term & synset.context | **00.52** | **56.30** | 37.46 | 44.99 | **47.34** |
| term & synset.context.hyp5 | **00.51** | **58.10** | **39.18** | **46.81** | **48.45** |
| term & synset.all | **00.52** | 56.19 | 37.44 | 44.94 | 47.32 |
| term & synset.all.hyp5 | **00.52** | **56.29** | **38.24** | **45.54** | 46.73 |

## Combine FCA & Text-clustering

1. preprocess Reuters documents and enrich them with background knowledge (Wordnet)
2. calculate a reasonable number *k* (100) of clusters with BiSec-*k*-Means using cosine similarity
3. extract a description for all clusters
4. relate clusters (objects) with FCA
5. use the visualization of the concept lattice for better understanding



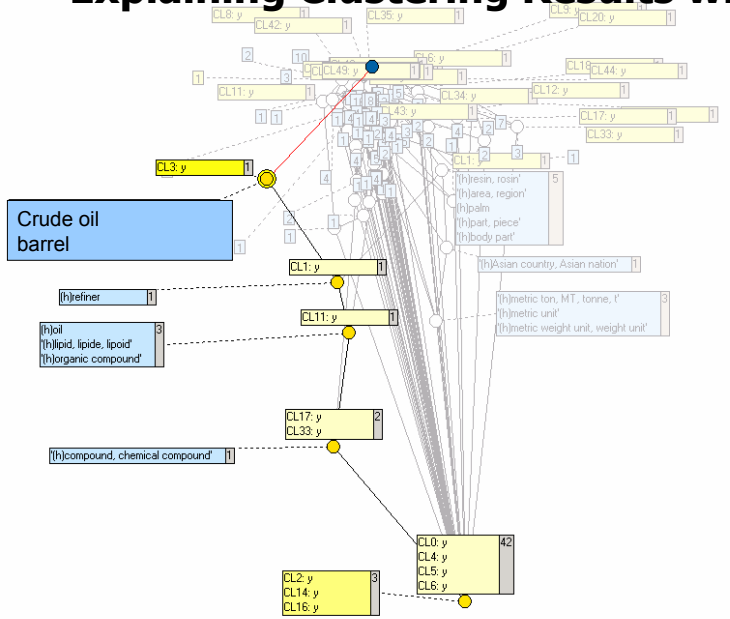
## Explaining Clustering Results with FCA



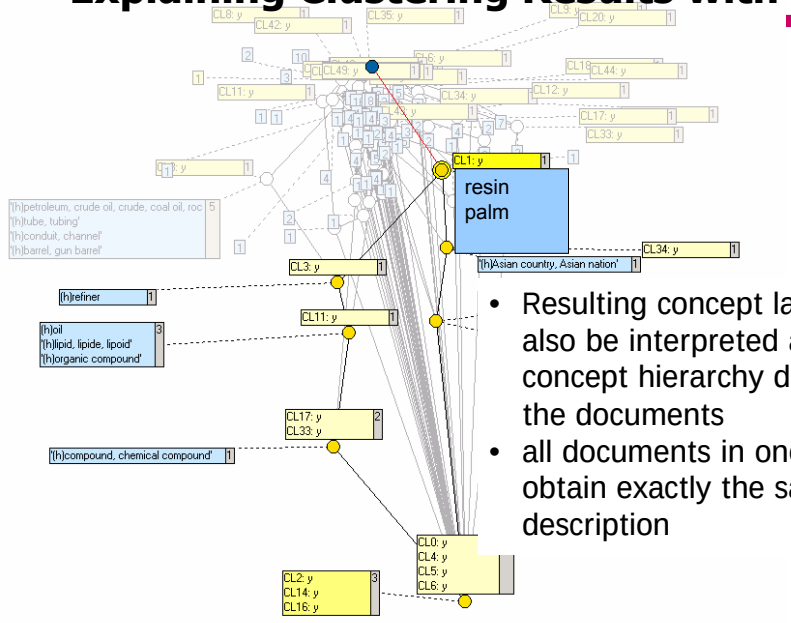

chain of concepts with increasing specificity



## Explaining Clustering Results with FCA



# Explaining Clustering Results with FCA




- Resulting concept lattice can also be interpreted as a concept hierarchy directly on the documents
- all documents in one cluster obtain exactly the same description



# Using Ontologies

- **Wordnet and IR**
  - Query expansion with wordnet does not really improve the performance

    Ellen M. Voorhees, Query expansion using lexical-semantic relations, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, p.61-69, July 03-06, 1994, Dublin, Ireland

- **Text Clustering and Ontologies**
  - Wordnet synset chains

    Green: Wordnet Chains (Stephen J. Green. Building hypertext links by computing semantic similarity. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 11(5):713–730, 1999.

  - Dave et.al.: worse results using an ontology (no word sense disambiguation)

    (Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the Twelfth International World Wide Web Conference, WWW2003*. ACM, 2003.)

  - Part of Speech attributes and named entities used as features

    (Vasileios Hatzivassiloglou, Luis Gravano, and Ankineedu Maganti. An investigation of linguistic features and clustering algorithms for topical document clustering. In *SIGIR 2000: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 24-28, 2000, Athens, Greece*. ACM, 2000.)



# Using Ontologies

## A kind of statistical concepts

- Calculating a kind of statistical concept and combine them with the classical bag of words representation

  L. Cai and T. Hofmann. Text Categorization by Boosting Automatically Extracted Concepts. In *Proc. of the 26th Annual Int. ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, Canada, 2003.

- Clustering word to setup a kind of concepts

  G. Karypis and E. Han. Fast supervised dimensionality reduction algorithm with applications to document categorization and retrieval. In *Proc. of 9th ACM International Conference on Information and Knowledge Management, CIKM-00*, pages 12–19, New York, US, 2000. ACM Press.

- Clustering words and documents simultaneously

  Inderjit S. Dhillon, Yuqiang Guan, and J. Kogan. Iterative clustering of high dimensional text data augmented by local search. In *2nd SIAM International Conference on Data Mining (Workshop on Clustering High-Dimensional Data and its Applications)*, 2002.



# Using Ontologies

## Text Classification and Ontologies

- Using Hypernyms of wordnet as concept feature (no WSD, no significant better results)

  Sam Scott , Stan Matwin, Feature Engineering for Text Classification, Proceedings of the Sixteenth International Conference on Machine Learning, p.379-388, June 27-30, 1999

- Brown Corpus tagged with Wordnet senses does not shows significant better results.

  A. Kehagias, V. Petridis, V. G. Kaburlasos, and P. Fragkou. A Comparison of Word- and Sense-Based Text Categorization Using Several Classification Algorithms. *Journal of Intelligent Information Systems*, 21(3):227–247, 2000.

- Map terms to concepts of the UMLS ontology to reduce the size of feature set, use search algorithm to find super concepts, evaluation using KNN and medline documents, show improvement.

  B. B. Wang, R. I. Mckay, H. A. Abbass, and M. Barlow. A comparative study for domain ontology guided feature extraction. In *Proceedings of the 26th Australian Computer Science Conference (ACSC-2003)*, pages 69–78. Australian Computer Society, 2003.

- Generative model consist of feature, concepts and topics, using Wordnet to initialize the parameter for concepts, evaluation on Reuter and Amazon corpus

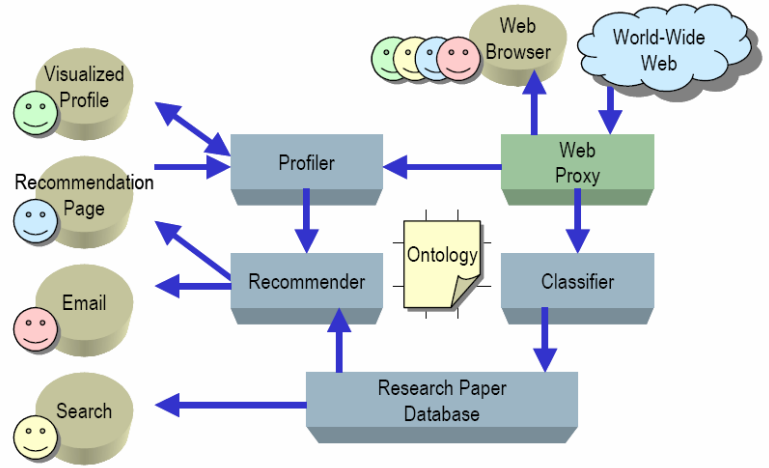
  Georgiana Ifrim, Martin Theobald, Gerhard Weikum, Learning Word-to-Concept Mappings for Automatic Text Classification Learning in Web Search Workshop 2005.

# Using Ontologies References


- Stephan Bloehdorn, Andreas Hotho: Text Classification by Boosting Weak Learners based on Terms and Concepts. ICDM 2004: 331-334
- Andreas Hotho, Steffen Staab, Gerd Stumme: Ontologies Improve Text Document Clustering. ICDM 2003: 541-544
- Andreas Hotho, Steffen Staab, Gerd Stumme: Explaining Text Clustering Results Using Semantic Structures. PKDD 2003: 217-228
- Stephan Bloehdorn, Philipp Cimiano, and Andreas Hotho: Learning Ontologies to Improve Text Clustering and Classification, Proc. of GfKl, to appear.




# Ontology-based Recommender System




(Middleton, Shadbolt 2004)



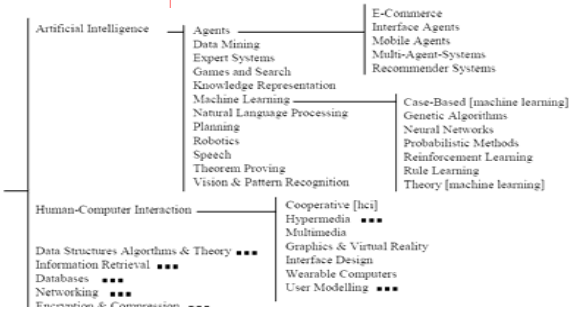
# Inferencing

$$\text{Topic interest} = \sum_{1..\text{no of instances}}^{n} \text{Interest value}(n) / \text{days old}(n)$$

| | |
|---|---|
| Event interest values | Paper browsed = 1<br>Recommendation followed = 2<br>Topic rated interesting = 10<br>Topic rated not interesting = -10 |
| Interest value for super-class per instance | = 50% of sub-class |

Improved recommendation accuracy

Less problems with cold start (user/System)




# Ontologies and Recommender References


- Middleton, S. E.; DeRoure, D.; and Shadbolt, N. R. 2003. Ontology-based recommender systems. In Staab, S., and Studer, R., eds., *Handbook on Ontologies*. Springer.
- Peter Haase, Marc Ehrig, Andreas Hotho, Björn Schnizler, Personalized Information Access in a Bibliographic Peer-to-Peer System, In Proceedings of the AAAI Workshop on Semantic Web Personalization, 2004, pp. 1-12. AAAI Press, July 2004.
- Peter Haase, Andreas Hotho, Lars Schmidt-Thieme, York Sure: Collaborative and Usage-Driven Evolution of Personal Ontologies. ESWC 2005: 486-499

# Agenda

- Introduction
- Foundations of the Semantic Web
- Ontology Learning
- Learning Ontology Mapping
- Semantic Annotation
- Using Ontologies
- Applications



# Application: Data Integration

- Data integration identified as $100Bs world-wide market
  - with significant govt interest creating a user-pull
    - Ontology development efforts, in OWL, aimed at information mgt ongoing in US govt include
      - NIST, NLM, EPA, DHS, DoD, DOJ, FDA, NIH, USGS, NOAA
- Huge potential follow-on market - EAI for the small business
  - making external data and info resources integrable
    - Could do for integration what Visicalc (excel) did for report generation



# Application: Ontoprise SemanticMiner

Company-wide Knowledge Management Project at Deutsche Telekom

**Goals**

- Make the Company's Competences
  - context
  - visible
  - usable
- Increase efficiency in sales and consulting

**Result**

- Integration of heterogeneous Sources
- Guided Search




# Why do KUKA Robotics apply Semantic Technologies

**Background**

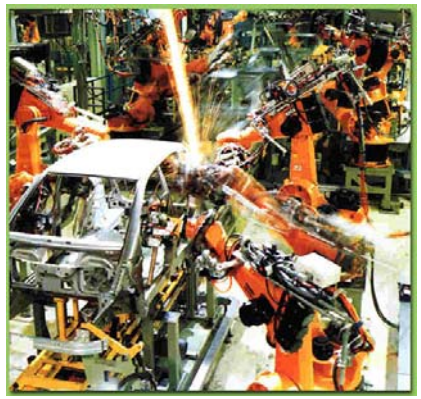
- 65% of all customer in the manufacturing industry change their suppliers because there are not satisfied with the service
- Service engineers spend a lot of time with known problems

**Goal**

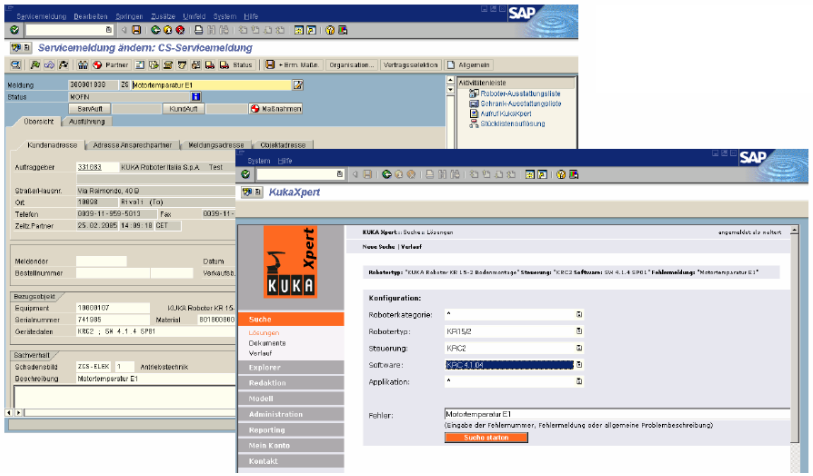
- Capturing and usage of engineers and experts know-how
- Decision support for choosing the right solution
- Increase customer satisfaction

**Implementation**

- Semantic Customer Service Support



# SemanticGuide: embedded in SAP CS & MAM





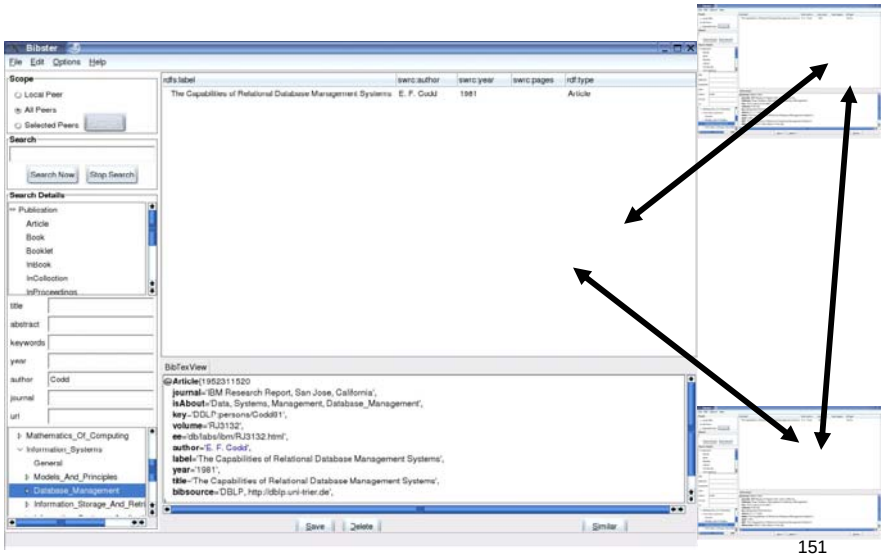

# Application: Web Services

Ultimate Goal: Application building by domain-experts rather than by software engineers
- Avoid expensive communication of knowledge
- Faster response to market needs

- Ontology Learning for Web Services: Creating Semantic Descriptions from other kind of structures (Sabou et al. WWW2005)
- Annotating Web Services by semantics
- Usage of both: Daniel Oberle „Semantic Management of Web Services“, Springer 2005/2006



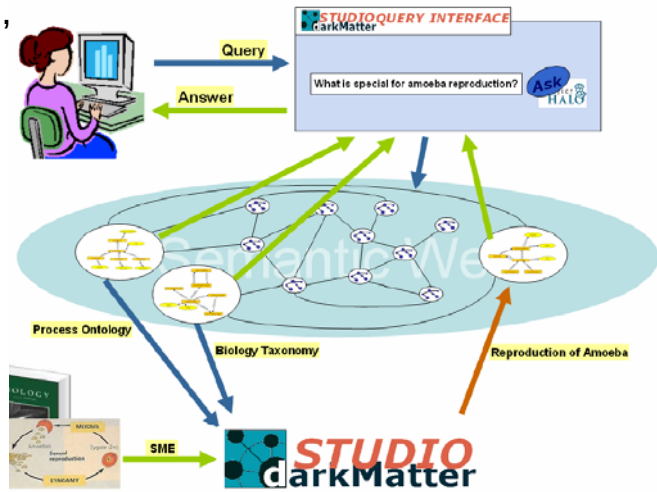
# Applications: Bibster






# Application: Project Halo

- Knowledge acquisition from textbooks
- Wikipedia like,
- for formal knowledge



# Application: Project Halo