Ontology Learning from Text

Paul Buitelaar, Philipp Cimiano, Marko Grobelnik, Michael Sintek

Tutorial at ECML/PKDD 2005

October 3rd, 2005

Porto, Portugal

In conjunction with the ECML/PKDD 2005 Workshop on:

Knowledge Discovery and Ontologies (KDO-2005)

Aims of the Tutorial

- Give an overview of Ontology Learning techniques as well as a synthesis of approaches
- Provide a ‘start kit’ for Ontology Learning
- Highlight interdisciplinary aspects and opportunities for a combination of techniques

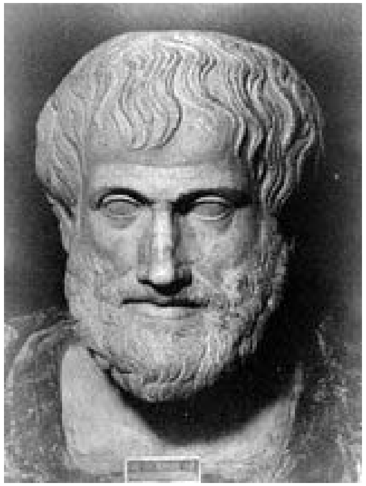
Structure of the Tutorial

Part I	**Introduction -** *Philipp Cimiano*
Part II	**Ontologies in Knowledge Management & Ontology Life Cycle -** *Michael Sintek*
Part III	**Methods in Ontology Learning from Text -** *Paul Buitelaar & Philipp Cimiano*
Part IV	**Ontology Evaluation -** *Marko Grobelnik*
Part V	**Tools for Ontology Learning from Text -** *All*
Wrap-up	*Paul Buitelaar*

Part I

Introduction to Ontologies and Ontology Learning

Aristotle - Ontology

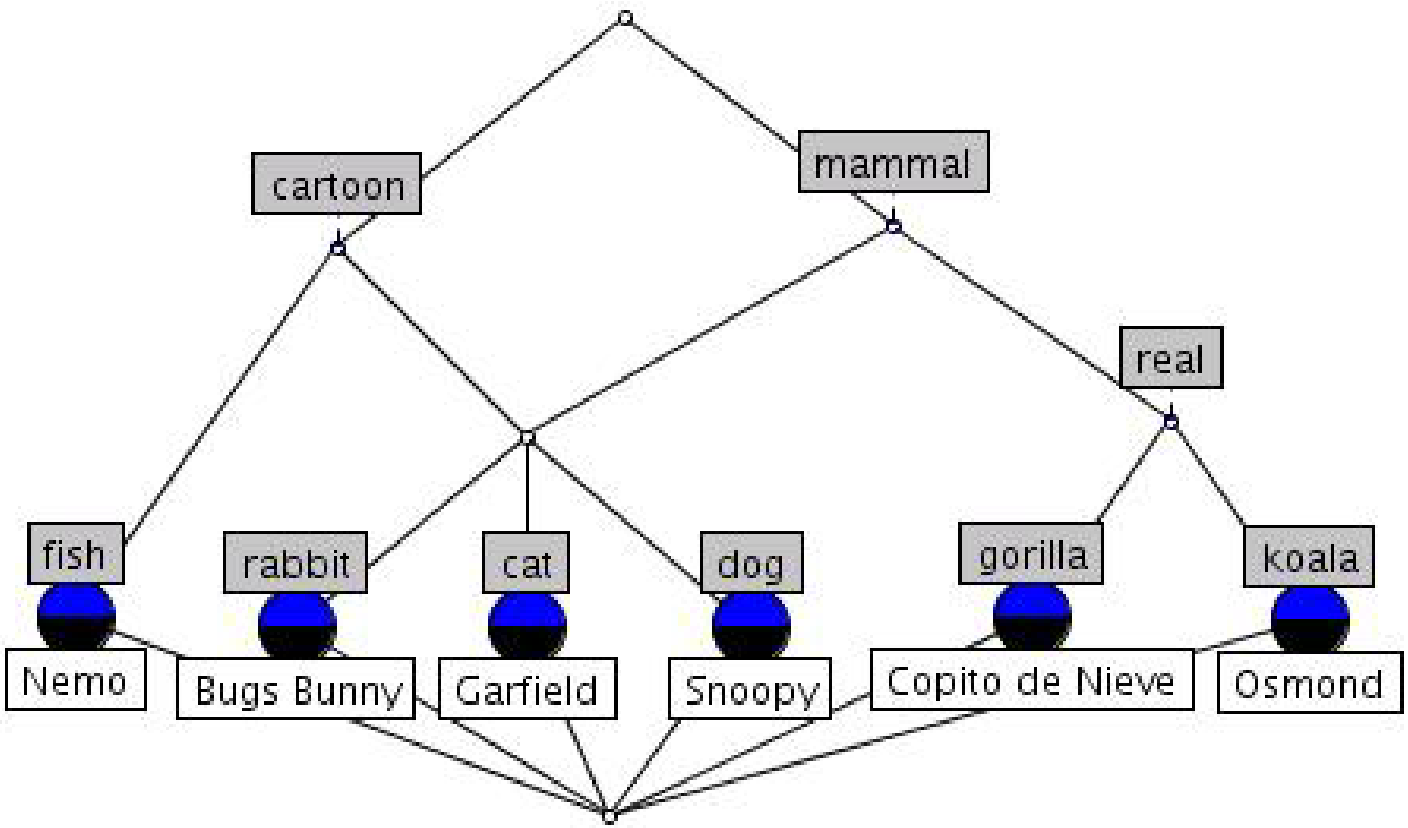


- Before: study of the nature of being
- Since Aristotle: study of knowledge representation and reasoning
- Terminology:
  - **Genus:** (Classes)
  - **Species:** (Subclasses)
  - **Differentiae:** (Characteristics which allow to group or distinguish objects from each other)
- Syllogisms (Inference Rules)

Example for differentiae (adapted from Uta Priss, in preparation)

	real	cartoon	cat	dog	rabbit	fish	gorilla	koala	mammal
Garfield		X	X						X
Snoopy		X		X					X
Bugs Bunny		X			X				X
Nemo		X				X			
Copito	X						X		X
Osmond	X							X	X

Organizing the Objects as a Lattice



Origin and History

- Ontology in Philosophy
  - a philosophical discipline, branch of philosophy that deals with the nature and the organization of reality

- Science of Being (Aristotle, Metaphysics, IV, 1)

- Tries to answer the questions:
  - *What characterizes being?*
  - *Eventually, what is being?*

Ontologies in Computer Science

- Ontology refers to an engineering artifact:
  - It is constituted by a specific vocabulary used to describe a certain reality, as well as
  - a set of explicit assumptions regarding the intended meaning of the vocabulary.

- An ontology is an explicit specification of a conceptualization. ([Gruber 93])
- An ontology is a shared understanding of some domain of interest. ([Uschold & Gruninger 96])

Why Develop an Ontology?

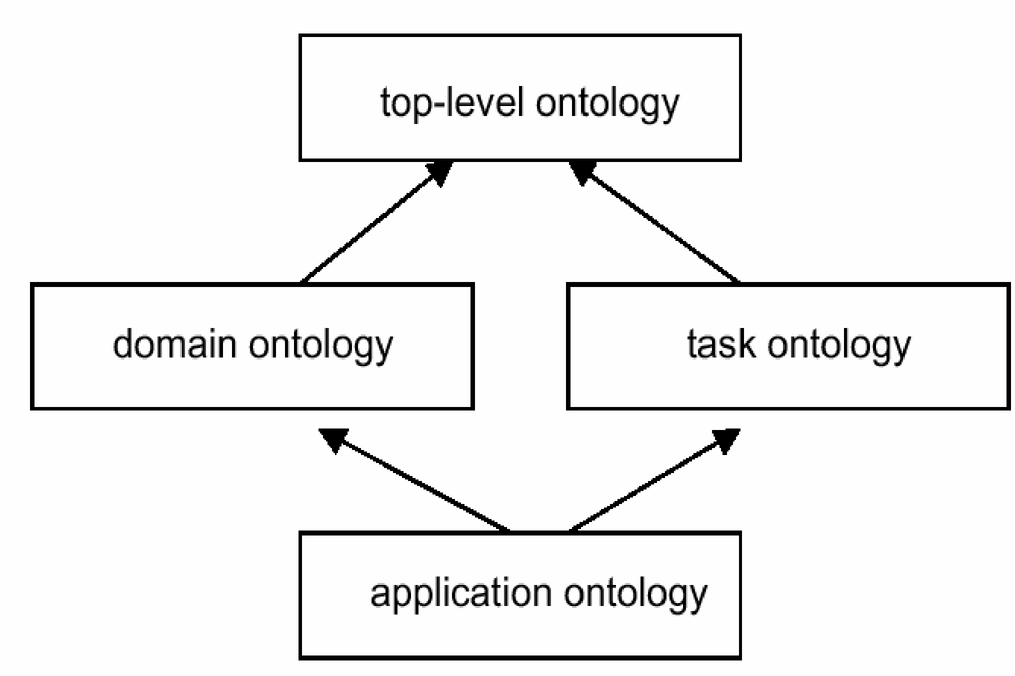
- To make domain assumptions **explicit**
  - Easier to change domain assumptions
  - Easier to understand and update legacy data
- To separate **domain knowledge** from operational knowledge
  - Re-use domain and operational knowledge separately
- A **community reference** for applications
- To **share a consistent understanding** of what information means

Types of Ontologies

[Guarino, 98]

Describe **very general concepts** like space, time, event, which are independent of a particular problem or domain. It seems reasonable to have unified top-level ontologies for large communities of users.

Describe the vocabulary related to a **generic domain** by specializing the concepts introduced in the top-level ontology.



Describe the vocabulary related to a **generic task or activity** by specializing the top-level ontologies.

These are the most specific ontologies. Concepts in application ontologies often correspond to **roles played by domain entities while performing a certain activity**.

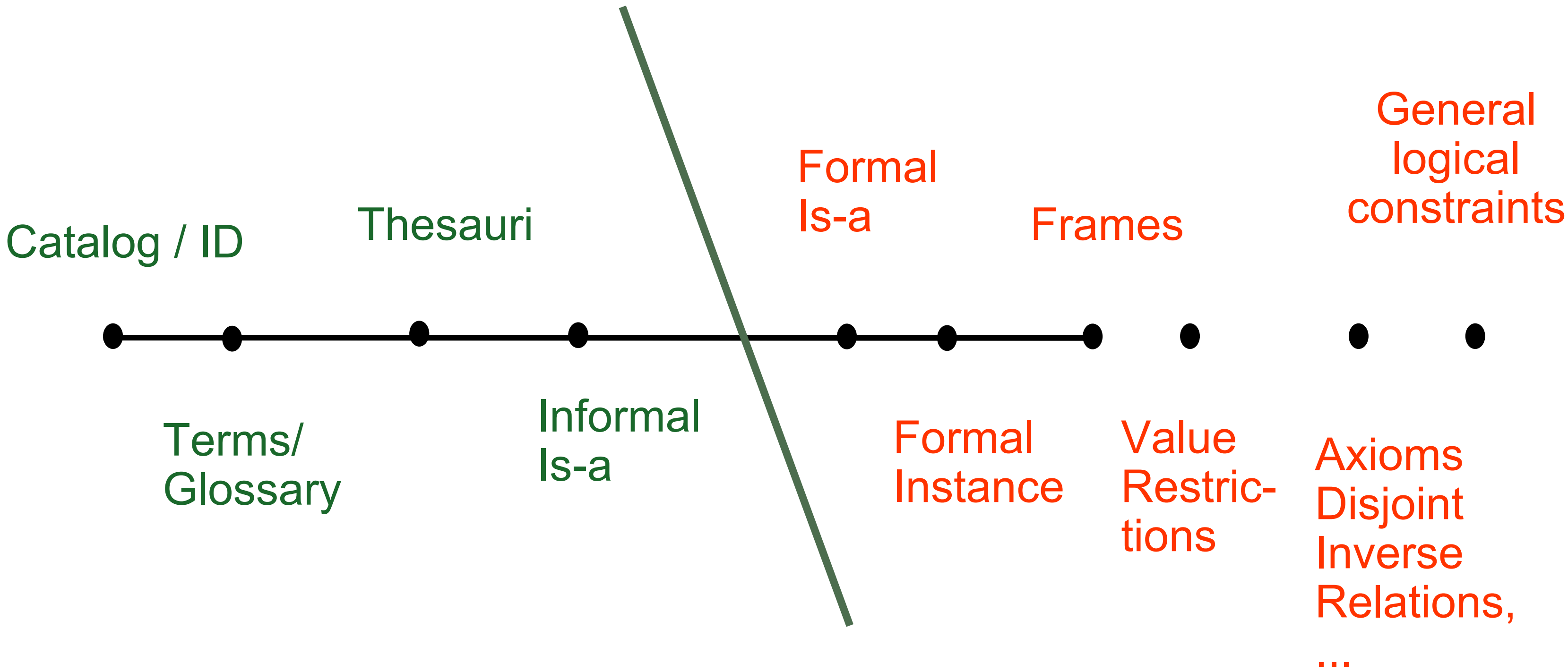
Ontologies - Some Examples

- **General purpose ontologies:**
  - WordNet, http://www.cogsci.princeton.edu/~wn
  - EuroWordNet

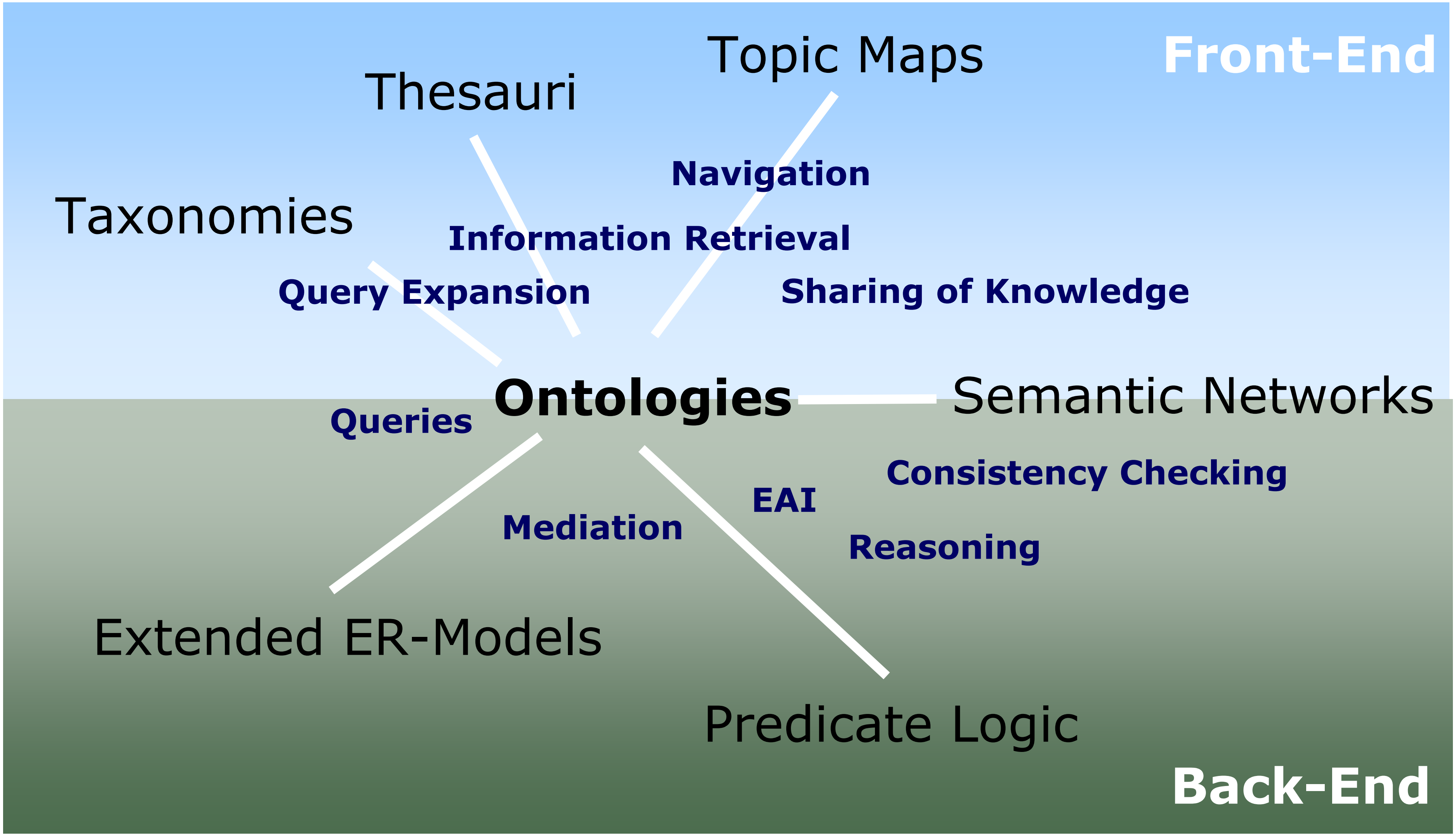
- **Upper level ontologies:**
  - DOLCE
  - Upper-Cyc Ontology, http://www.cyc.com/cyc-2-1/index.html
  - IEEE Standard Upper Ontology, http://suo.ieee.org/

- **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/
  - RETSINA Calendering Agent, http://ilrt.org/discovery/2001/06/schemas/ical-full/hybrid.rdf
  - AIFB Web Page Ontology, http://ontobroker.semanticweb.org/ontos/aifb.html
  - Web-KB Ontology, http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wwkb/
  - Dublin Core, http://dublincore.org/

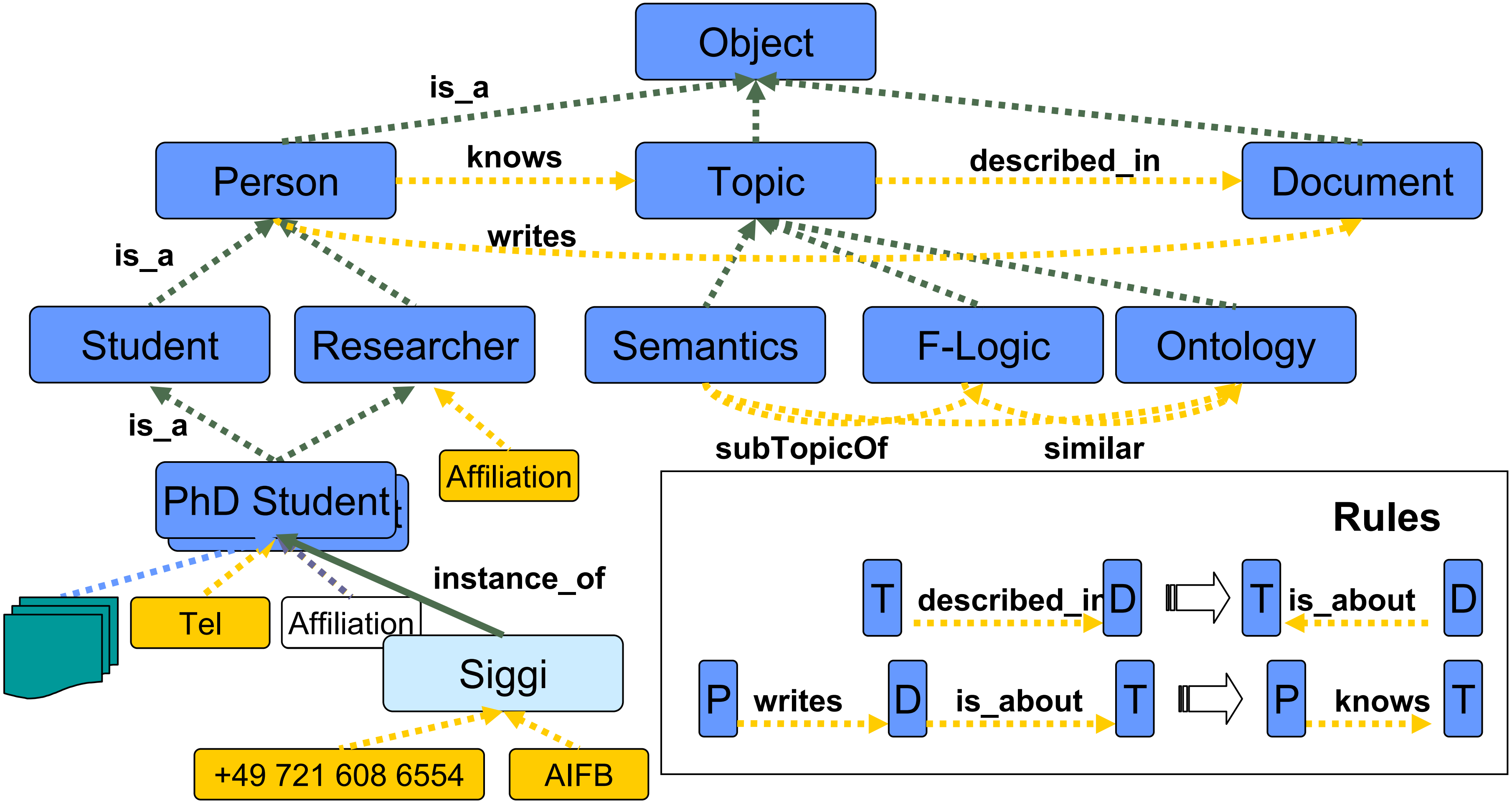
Ontologies and Their Relatives



Ontologies and Their Relatives (cont´d)



Ontology (in our sense)



Applications of Ontologies (adapted from [Sure 2003])

- **Natural Language Processing and Machine Translation**, e.g. Nirenburg et al. 2004, Maedche et al. 2001, Agirre et al. 1996, Beale et al. 1995
- **Semantic Web,** see http://www.w3.org/2001/sw/ and http://www.w3.org/2001/sw/WebOnt/
- **Knowledge Engineering & Management**, e.g. Fensel 2001, Mullholland et al. 2000; Staab & Schnurr, 2000; Sure et al., 2000, Abecker et al. 1997
- **Electronic Commerce**, e.g. RosettaNet3 and Ontology.org4
- **Information Retrieval and Information Integration**, e.g. Kashyap, 1999; Mena et al., 1998; Wiederhold, 1992
- **Intelligent Search Engines**, e.g. WebKB (Martin et al. 2000), SHOE (Heflin & Hendler, 2000), OntoSeek (Guarino et al., 1999), Ontobroker (Decker et al., 1999)
- **Digital Libraries**, e.g. Amann & Fundulaki, 1999
- **Enhanced User Interfaces**, e.g. (Kesseler, 1996), Inxight5
- **Software Agents**, e.g. OnTo-agents, FIPA, (Gluschko et al., 1999; Smith & Poulter, 1999)
- **Business Process Modeling**, e.g. Decker et al., 1997; TOVE, 1995; Uschold et al., 1998

The Mathematical Definition of an Ontology

[Stumme et al.]

- Structure: $\mathcal{C} := (C, <_C, R, <_R, \sigma)$
  - C: set of concept identifiers
  - R: set of relation identifiers
  - $<_C$ partial order on C (concept hierarchy)
  - $<_R$: partial order on R (relation hierarchy)
  - Signature: $\sigma : R \rightarrow C^+$

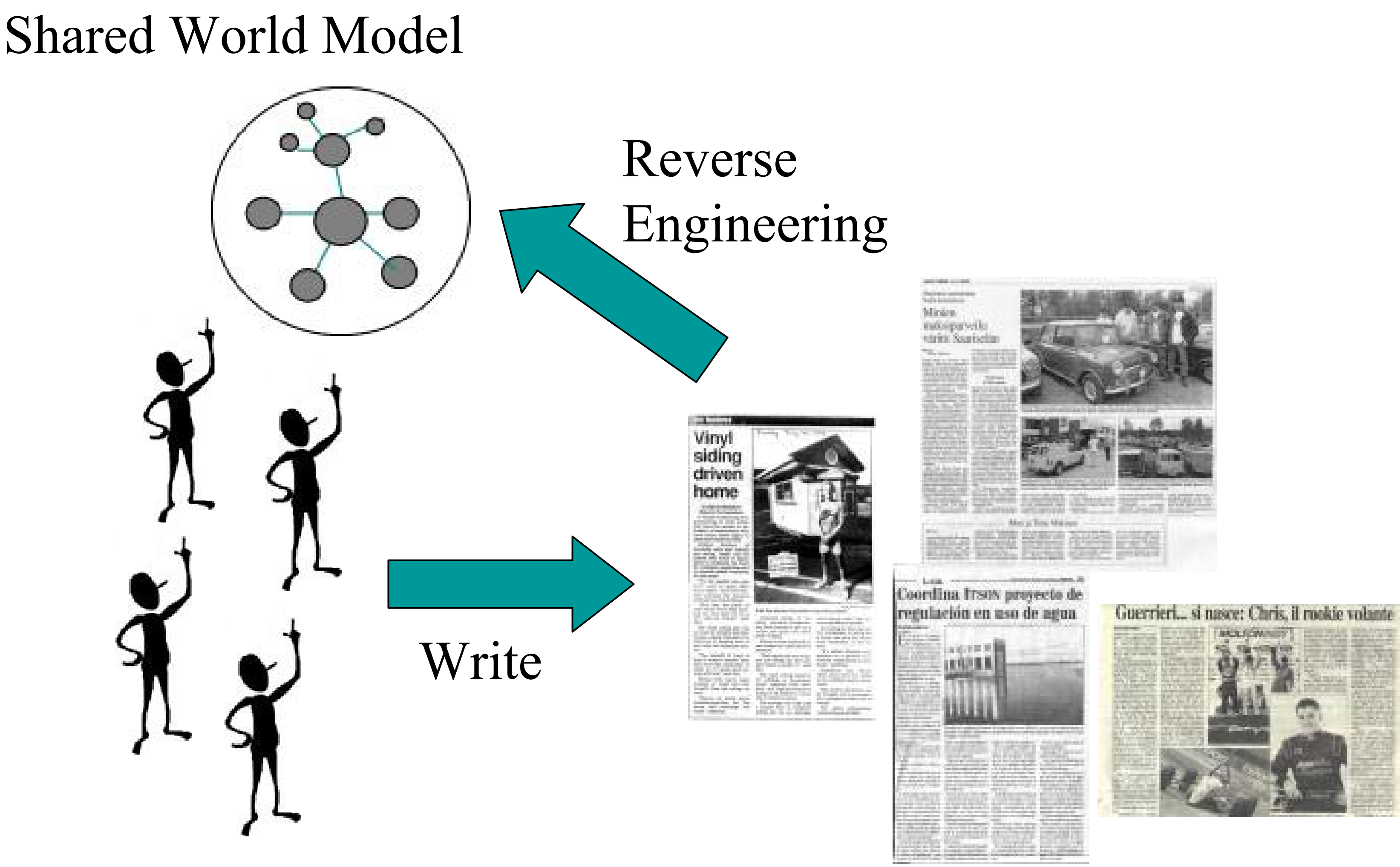
  - Mathematical definition of extension of concepts [c] and relations [r]

  - L-Axiom System: $\alpha(\text{disjoint}(\text{c}, \text{c}')) = \forall e \in [c] \rightarrow e' \notin [c']$

Motivation for Ontology Learning from Text

- **Problem**:
  - Knowledge Acquisition Bottleneck
- **Possible solution**:
  - Data-driven Knowledge Acquisition
  - As text is massively available on the Web, ontology learning from text is an attractive option

OL from Text as Reverse Engineering



Ontology Learning Layer Cake

$\forall x, y\,(sufferFrom\,(x, y) \rightarrow ill(x))$	Axioms & Rules
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Introduced in: Philipp Cimiano, PhD Thesis University of Karlsruhe, forthcoming

Part II

Ontologies in Knowledge Management & Ontology Life Cycle

Ontologies in Knowledge Management

Mainly based on work at DFKI Knowledge Management Department, Kaiserslautern

Knowledge Management (KM) and Ontology Learning

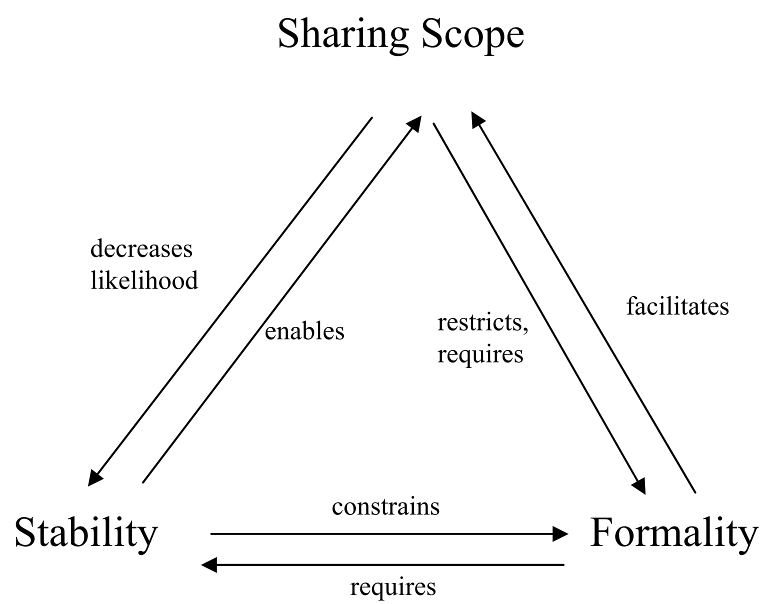
- KM is one of the main areas for ontology use and therefore gives input for various ontology learning aspects
- Well-established knowledge life cycle inspires ontology life cycle (→ ontology evolution/ management/negotiation) with ontology learning as important component

Ontologies in Information Systems for Knowledge Management

- **Idea: Shared vocabulary (concepts, relations, axioms) of the various actors in a KM information system**
- **Scientific questions:**
  - *Creation and maintenance, goal* “use time” >> “formalization time”
    - Which representation (taxonomy, frame logic, description logic)
    - Which concepts, relations, axioms (*conceptualization)*
    - How are they established between actors (sharing, semi-automatically) → **ontology learning!**
  - Usage for
    - Information presentation (personal views)
    - Retrieval
    - Information extraction
    - Reasoning
    - Knowledge conservation

Degree of Formality Interacts with Sharing Scope and Stability of Knowledge

- Formalization is expensive in terms of time and money
  - requires:
    „use time“ >> „formalization time“
    i.e., high stability required
  - but: stability mostly externally given
- Formality allows for sharing (explicitness, precision)
  - prerequisites formal training
  - possibly keeps away agents from participation
  - wide sharing scope increases costs of negotiation

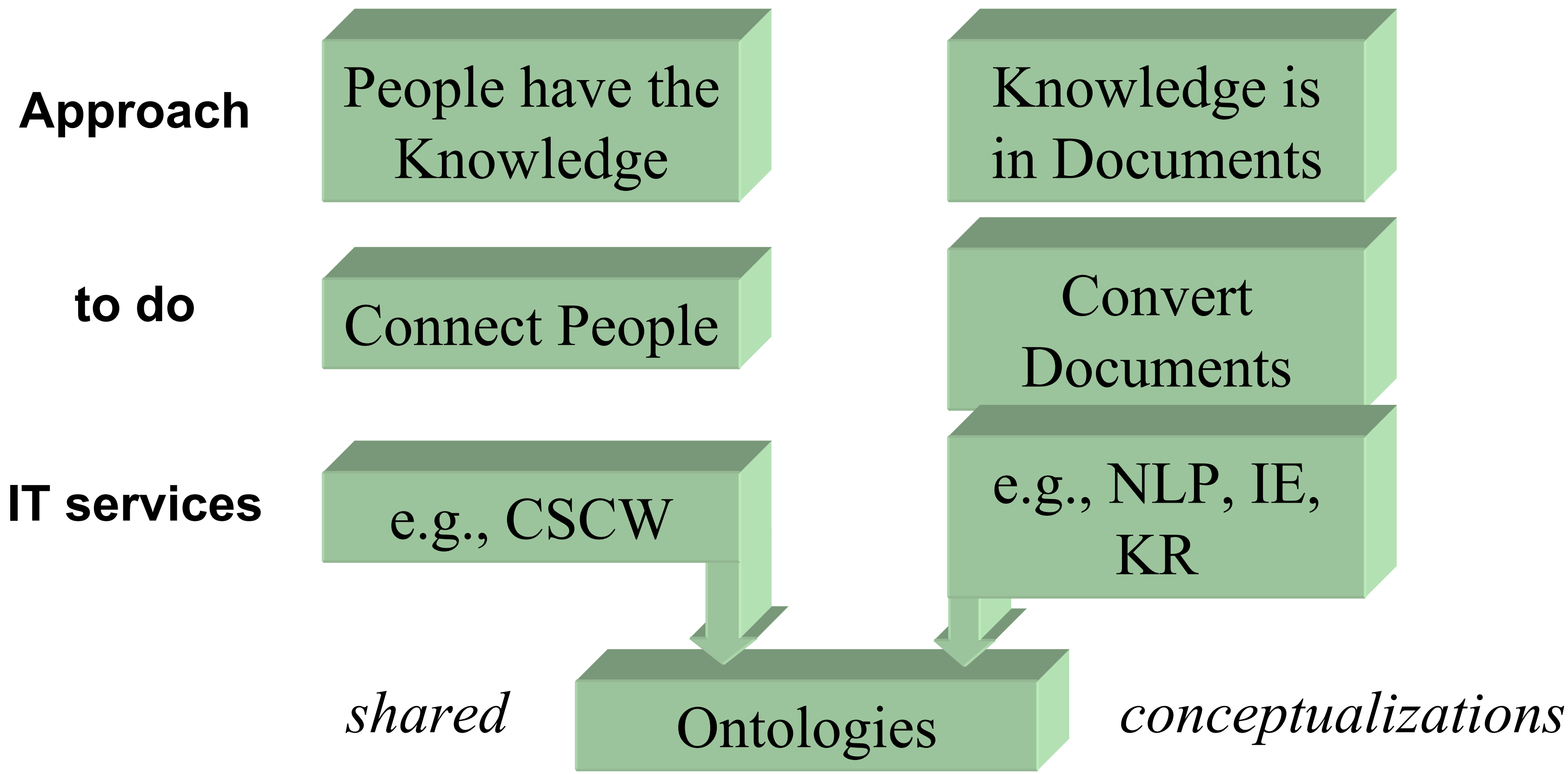


Ontology Management and Negotiation

- Ontology Management is an important means to *balance between local and global concerns* in Distributed Organizational Memory scenarios

- Ontology Negotiation needs (at least)
  - Overlap detection and evidence integration
  - Negotiation speech acts and protocols
  - Explicit handling of the sharing scope (societies)

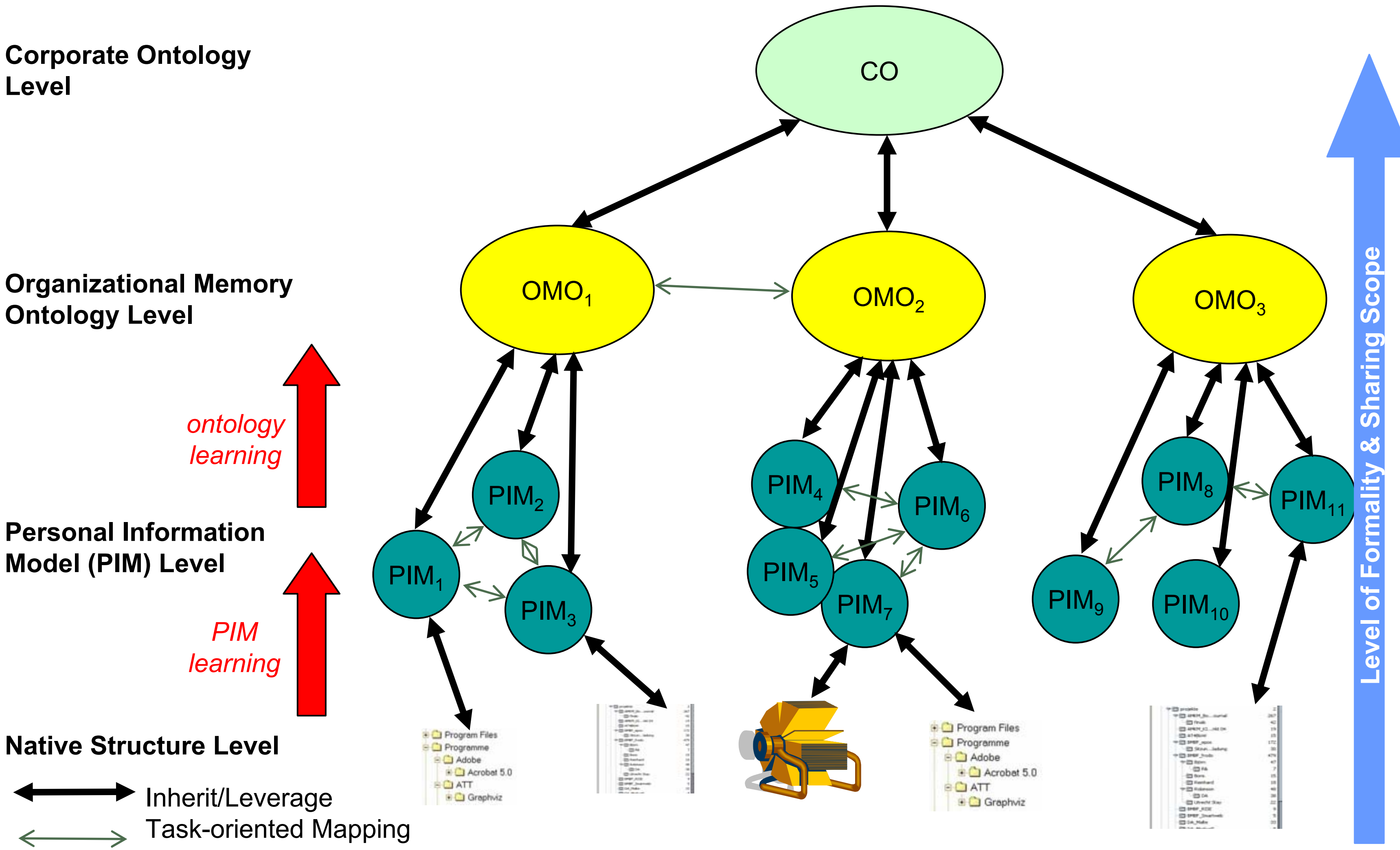
Ontologies Span Two Lines of Action in KM



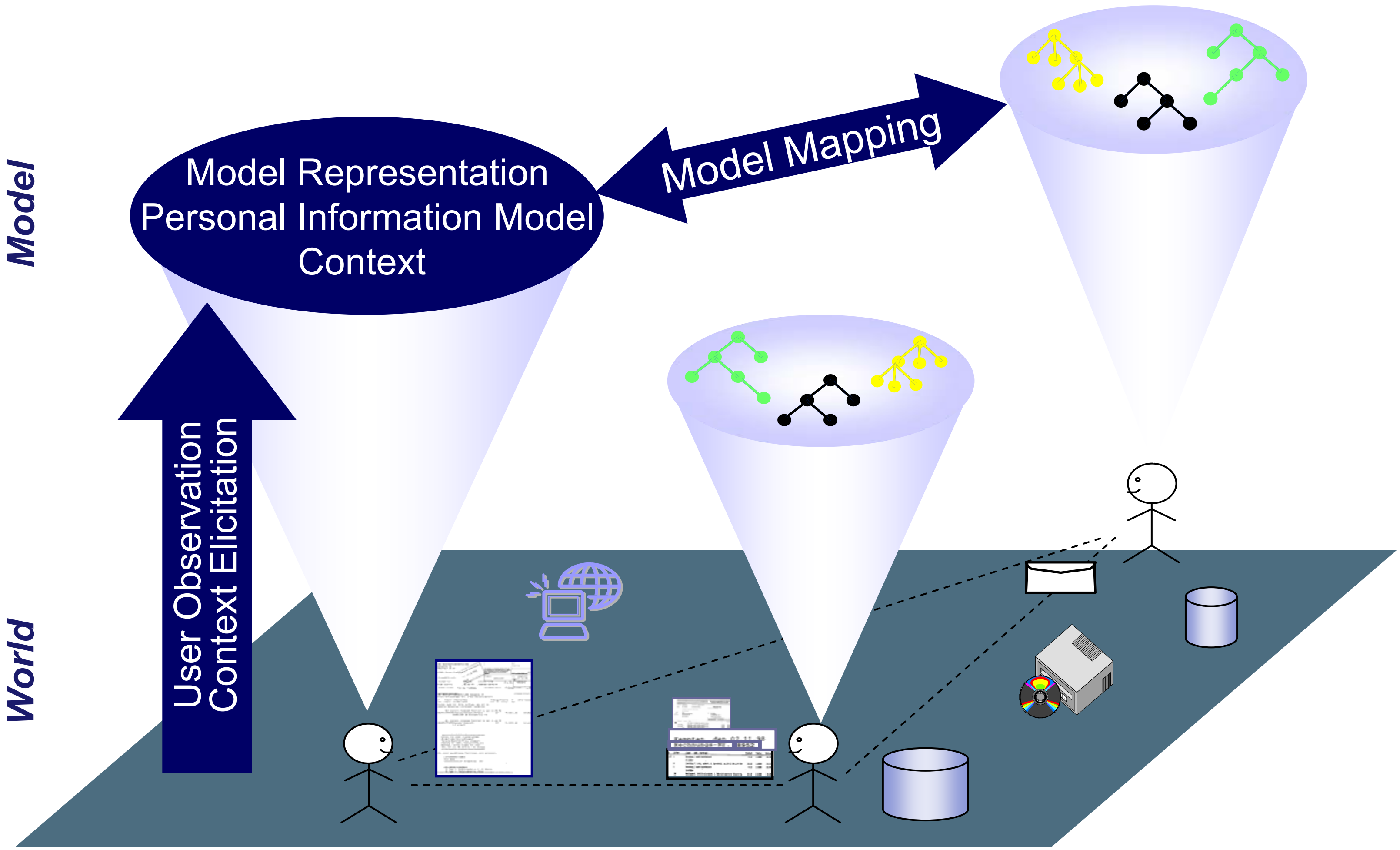
Personal Information Models vs. Ontologies

- In KM, we distinguish between *personal* information models and "*shared*" ontologies
- The personal information model is a formally grounded model reflecting aspects of a *knowledge worker's view* on his information landscape
- More global ontologies as well as native structures provide input for personal information models, and personal information models provide input for more global ontologies
- The personal information model can be utilized by various knowledge services (retrieval, personal information agent, visualization, …)
- Research Topics:
  - Leveraging native structures (file folders, e-mail folders, address book entries, mind maps, personal wikis; supported by *documents* in these structures…)
  - Integration of/into existing ontologies
  - Mappings between personal information models
  - → Learning of personal information models as basis for ontology learning

Ontology Space (EPOS Project)

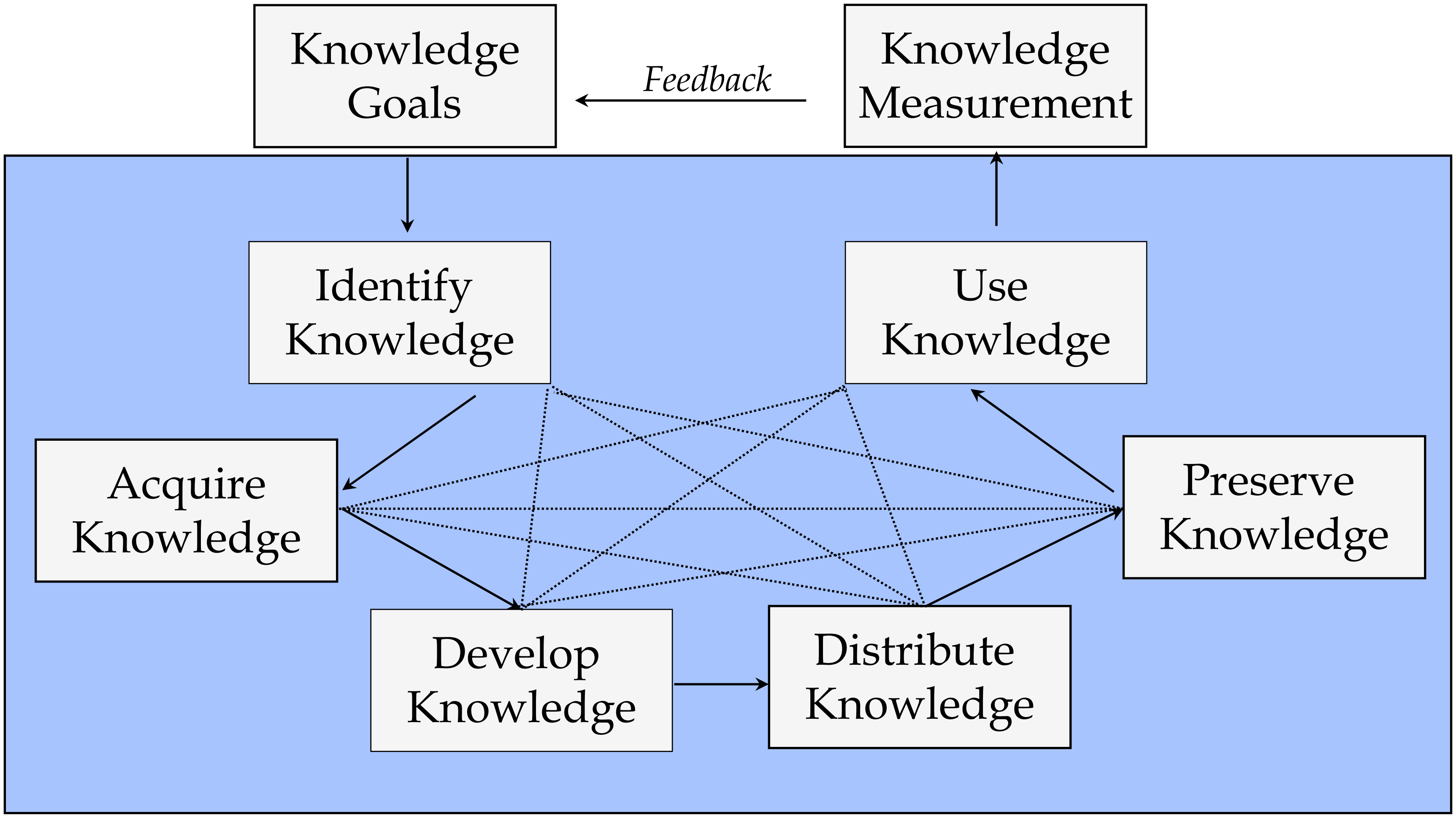


Representation, Acquisition, and Mapping of Personal Information Models is at the heart of KM Research



Ontology Life Cycle

Building Blocks for Knowledge Management Processes I



Adapted from: Probst/Raub/Romhardt

Building Blocks for KM Processes II

- *Knowledge Goals*
  - point the way for knowledge management activities
  - can be normative, strategic, or operational
- *Knowledge Identification*
  - companies should know what knowledge and expertise exist both inside and outside their own walls
  - most big companies lose track of their internal and external data, information, and capabilities.
- *Knowledge Acquisition*
  - Knowledge can be acquired via the following “import channels”: (1) Knowledge Held by Other Firms; (2) Stakeholder Knowledge; (3) Experts; (4) Knowledge Products
- *Knowledge Development*
  - Knowledge development consists of all the management activities intended to produce new internal or external knowledge on both the individual and the collective level

Building Blocks for KM Processes III

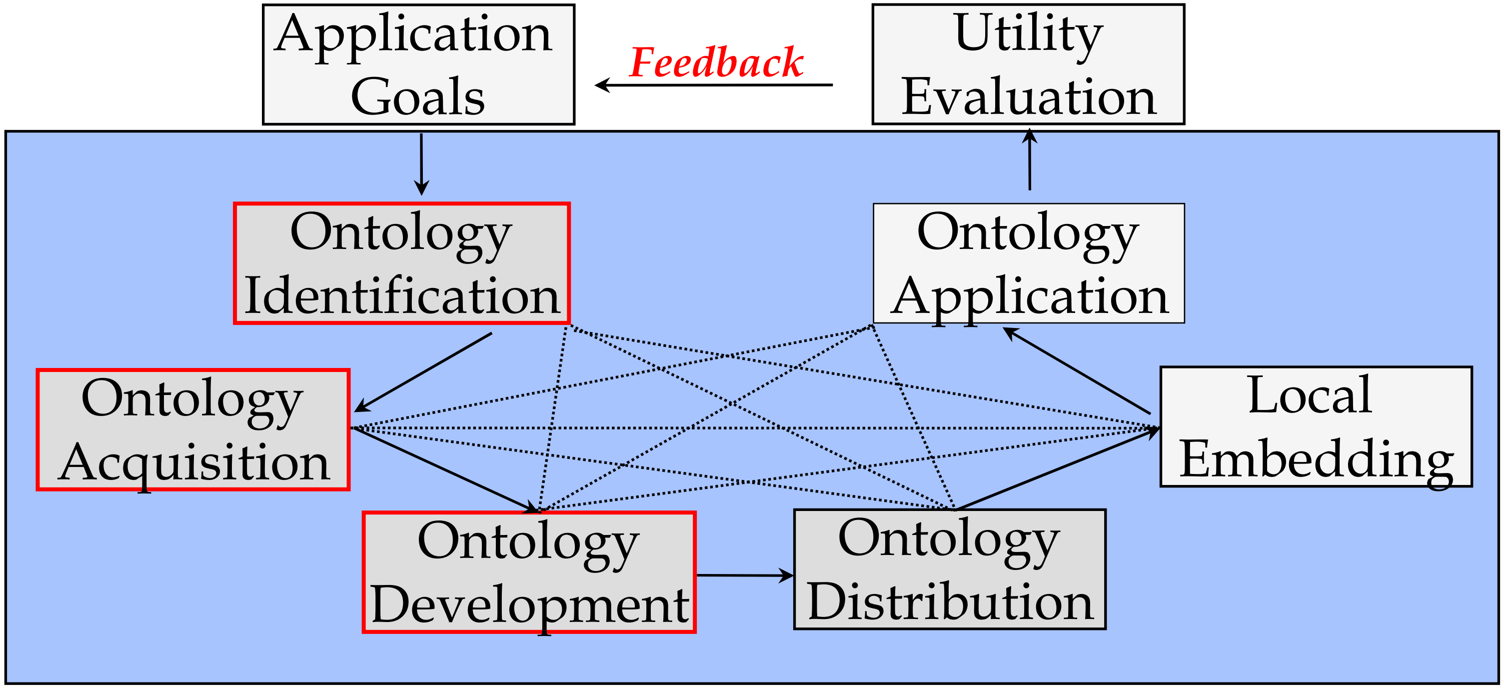
- *Knowledge Distribution*
  - make knowledge available and usable across the whole organization
  - critical questions: Who should know what, to what level of detail, and how can the organization support these processes of knowledge distribution?
  - Relevant technologies: groupware, modern forms of interactive management information systems, and all instruments of computer-supported cooperative work

- *Knowledge Preservation*
  - After knowledge has been acquired or developed, it must be carefully preserved
  - To avoid the loss of valuable expertise, companies must shape the processes of selecting valuable knowledge for preservation, ensuring its suitable storage, and regularly incorporating it into the knowledge base

- *Knowledge Use*
  - productive deployment of organizational knowledge in the production process
  - is the purpose of knowledge management

- *Knowledge Measurement*
  - biggest challenge in the field of knowledge management: no tested tool box of accepted indicators and measurement processes
  - knowledge and capabilities can rarely be tracked to a single influencing variable
  - cost of measuring knowledge is often seen as too high

Ontology Life Cycle Analogous to KM Life Cycle



"Relevant for OL in RED"

- *Ontology identification and acquisition* are triggered from application use, documents and from feedback from the previous loop
- Ontologies are *locally embedded* in the concrete usage context; this is necessary since usual not all parts of an ontology are useful in a certain context (like manufacturing aspects for the bookkeeping applications)

Consequences from Ontology Life Cycle for Ontology Learning

- Feedback:
  - Not only explicit feedback (semi-automatic OL), but also implicit (feedback wrt. application goals)
- Support of Ontology Evolution & Versioning
  - Change management
  - Inconsistency management
- Ontology Evaluation (Part IV)

Ontology Evolution – Requirements

- **Functionality**
  - enable the handling of ontology **changes**
  - ensure the **consistency** of the underlying ontology and all dependent artifacts, e.g., instances
- **Guiding the user**
  - support the user to manage changes **more easily**
- **Refining the ontology**
  - offer advice to the user for **continual** ontology refinement
  - discover changes that lead to an **improved** ontology

From: Studer & Haase

Representation of Proposed Ontology Changes

- Syntactic and algebraic
  - Ontology algebras (cf. Wiederhold):
  - Operations: intersection, union, difference

- Semantic
  - Based on model theory (cf. Sintek et al., 2004 “A Formalization of Ontology Learning from Text”)
  - Operations do not take (syntactical) ontology representation into account, but their semantics
  - Necessary for complex ontology languages like OWL

Ontology Change Operators + and – : Ontology entailment

- we define the semantics of $+$ and $-$ with the help of ontology *entailment*
- an ontology O_1 is said to *entail* an ontology O_2, written $O_1 \models O_2$, if every model for O_1 is also a model for O_2
- an ontology O is said to be a *most general* ontology for a condition C if O fulfills C and there exists no other ontology $O' \neq O$ which fulfills C and for which $O \models O'$ holds
- in general, more than one most general ontology for a condition exists
- a *least general* ontology for a given condition is specified analogously

From: Michael Sintek et al., 2004 "A Formalization of Ontology Learning from Text"

Definition of + and –

- $O_1 + O_2$ is a most general ontology O with $O \models O_1 \;\wedge O \models O_2$
- $O_1 - O_2$ is a least general ontology O with $O_1 \models O \;\wedge O \not\models O_2$
- in general, the result of $O_1 + O_2$ and $O_1 - O_2$ is not well-defined, depending on the choice of the ontology language
- $+$ and $-$ are not symmetric: $+$ adds *all* of O_2 to O_1, while $-$ removes only a minimal portion of O_2 from O_1
- $+$ is usually well-defined since most ontology languages allow the statements that are used to define ontologies to be joined in some simple way
- $-$ is usually not well-defined, thus leaving several choices to the user; this does, however, not cause any problems in our scenario since the user has to interact with the suggestions anyway

Example Usage (From OntoLT System)

- rules for turning the subject of a sentence to a class and the predicate to a slot with the direct object as range:

  $\forall S$ subject$(S) \leftarrow$

  xpath$(\Sigma,$ " `.//phrase[@id=...[@type="SUBJ"]/@phrase]/head`", $S) \wedge \ \ldots .$

  $\forall P$ predicate$(P) \leftarrow$

  xpath$(\Sigma,$ " `.//phrase[@id=./../..//clause/@pred]`", $P)$ $\wedge \ \ldots .$

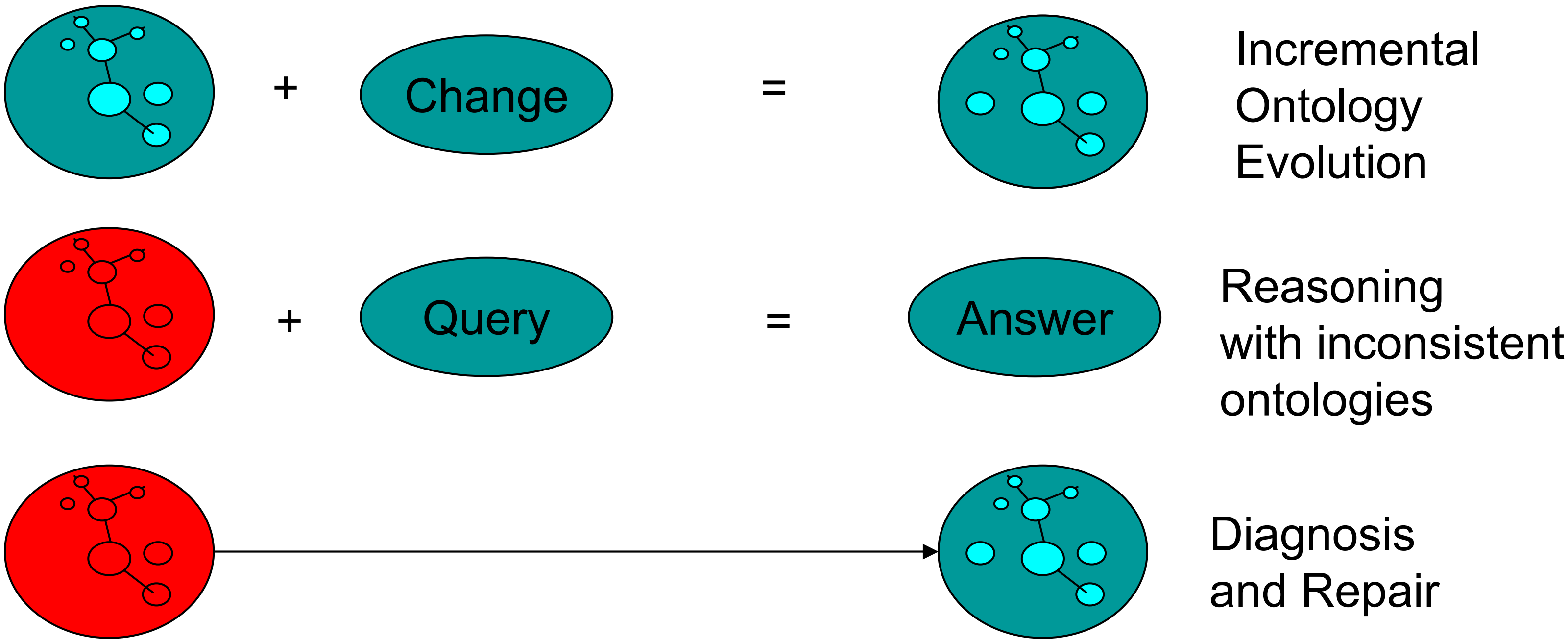
  $\forall O$ directObject$(O) \leftarrow$

  xpath$(\Sigma,$ " `.//phrase[@id=...[@type="DOBJ"]/@phrase]/head`", $O) \wedge \ \ldots .$

  $\forall S, P, O$ subject$(S) \wedge$ predicate$(P) \wedge$ directObject$(O) \mapsto$

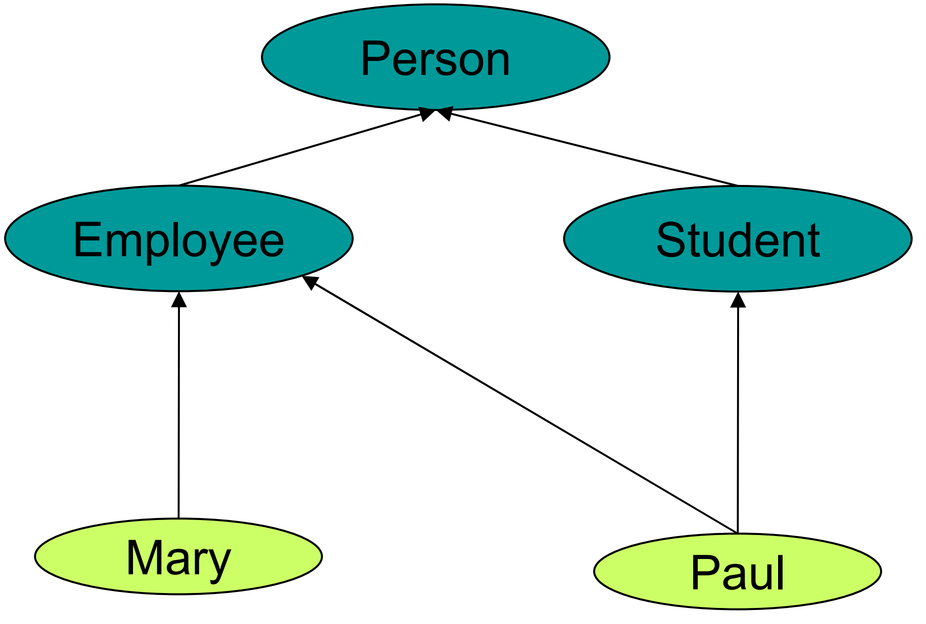
  $+\{S \sqsubseteq \top, O \sqsubseteq \top, S \sqsubseteq \forall P.O\}.$

Approaches for Inconsistency Management



From: Studer & Haase

Sample Ontology



Student $\subseteq$ Person
Employee $\subseteq$ Person
Employee(Mary)
Employee(Paul)
Student(Paul)

Logical Consistency

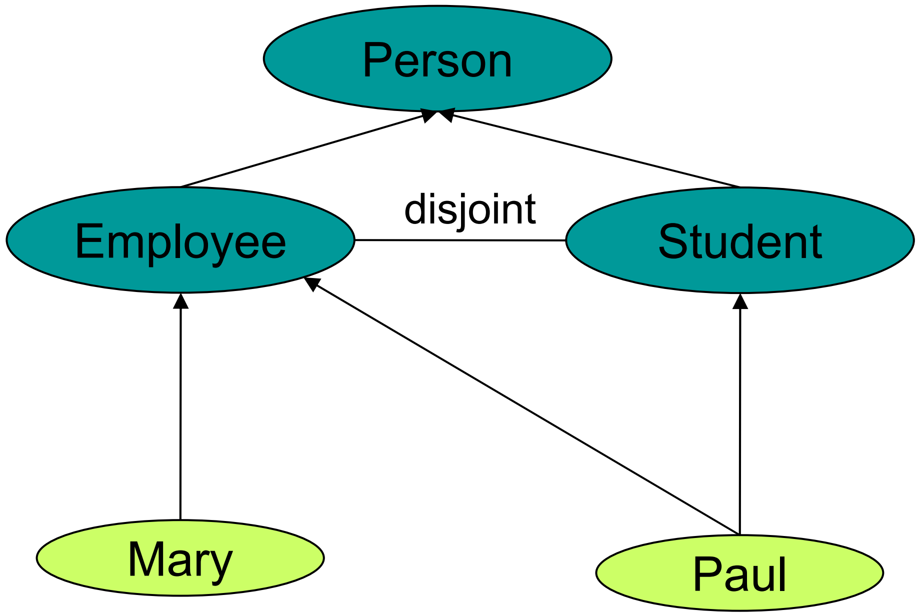
- Consistency condition: ontology must be satisfiable, i.e. it must have a non-empty model

- Why is this important?
  - An inconsistent ontology entails every fact:

    KB |= α for every α

  - Query answering would become meaningless!

Logical Consistency



- Ontology has no model, i.e., is logically inconsistent
- Resolution Function: Alternatives
  - Find a minimal inconsistent sub-ontology
  - Find a maximal consistent sub-ontology

Part III

Methods in Ontology Learning from Text

Some pre-History

- AI: Knowledge Acquisition
  - Since 60s/70s: Semantic Network Extraction and similar for Story Understanding
    - Systems: e.g. MARGIE (Schank et al., 1973), LUNAR (Woods, 1973)
- NLP: Lexical Knowledge Extraction
  - 70s/80s: Extraction of Lexical Semantic Representations from Machine Readable Dictionaries
    - Systems: e.g. ACQUILEX LKB (Copestake et al.)
  - 80s/90s: Extraction of Semantic Lexicons from Corpora for Information Extraction Systems
    - Systems: e.g. AutoSlog (Riloff, 1993), CRYSTAL (Soderland et al., 1995)
- IR: Thesaurus Extraction
  - Since 60s: Extraction of Keywords, Thesauri and Controlled Vocabularies
    - Based on construction and use of thesauri in IR (Sparck-Jones, 1966/1986, 1971)
    - Systems: e.g. Sextant (Grefenstette, 1992), DR-Link (Liddy, 1994)

Some Current Work on Ontology Learning from Text

Term Extraction
- Statistical Analysis
- Patterns
- (Shallow) Linguistic Parsing
- Term Disambiguation & Compositional Interpretation
- Combinations

Taxonomy Extraction
- Statistical Analysis & Clustering (e.g. FCA)
- Patterns
- (Shallow) Linguistic Parsing
- WordNet
- Combinations

Relation Extraction
- Anonymous Relations (e.g. with Association Rules)
- Named Relations (Linguistic Parsing)
- (Linguistic) Compound Analysis
- Web Mining, Social Network Analysis
- Combinations

Relation Label Extraction
- Extension of Association Rules Algorithm

Definition Extraction
- (Linguistic) Compound Analysis (incl. WordNet)

Some Current Work on Ontology Learning from Text

AIFB – *TextToOnto* (Maedche and Staab, 2000; Cimiano et al., 2005)

- Term Extraction and Taxonomy Extraction
  - Statistical Analysis
  - Conceptual Clustering (FCA), Patterns, WordNet (+ Combination)
- Relation Extraction
  - Anonymous Relations (Associaton Rules)
  - Named Relations (Subcategorization Frames)

CNTS Univ. Antwerpen, VUB (Reinberger et al., 2004)

- Concept Formation + Relation Extraction
  - Shallow Linguistic Parsing
  - Clustering

DFKI – *OntoLT* (Buitelaar et al., 2004), *RelExt* (Schutz and Buitelaar, 2005)

- Term Extraction
  - Shallow Linguistic Parsing & Statistical Analysis
- Taxonomy and Relation Extraction
  - Shallow Linguistic Parsing & manually defined mapping rules
  - Named Relations (Subcategorization Frames)

Some Current Work on Ontology Learning from Text

Economic Univ., Prague (Kavalec and Svatek, 2005)

- Relation Label Extraction
  - Extension of Association Rules Algorithm

Free Univ. Amsterdam (Sabou, 2005)

- Term and Taxonomy Extraction (for Web Service Ontologies)
  - Shallow Linguistic Analysis & Patterns

Jozef Stefan Inst., Ljubljana -- *OntoGen* (Fortuna et al., 2005)

- Term and Taxonomy Extraction
  - Statistical Analysis & Clustering
- Relations
  - Web Mining, Social Network Analysis

Univ. Paris -- *ASIUM* (Faure and Nedellec, 1998)

- Taxonomy Extraction (& Subcategorization Frames)
  - Shallow Linguistic Parsing
  - Clustering

Some Current Work on Ontology Learning from Text

Univ. Rome – *OntoLearn* (Navigli and Velardi, 2004; Velardi et al., 2005)

- Term Extraction and Interpretation
  - Shallow Linguistic Parsing &Term Disambiguation & Compositional Interpretation
- Relations
  - Classification of the relation between terms in a compound into predefined set of (thematic) relations
- Definitions
  - Rules for Gloss Generation

Univ. of Zürich (Rinaldi et al., 2005)

- Term and Taxonomy Extraction
  - Shallow Linguistic Analysis & Patterns

Overview of Current Work: Paul Buitelaar, Philipp Cimiano, Bernardo Magnini *Ontology Learning from Text: Methods, Evaluation and Applications* Frontiers in Artificial Intelligence and Applications Series, Vol. 123, IOS Press, July 2005.

Ontology Learning Layer Cake

Example	Layer
$\forall x, y\,(sufferFrom(x, y) \rightarrow ill(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Introduced in: Philipp Cimiano, PhD Thesis University of Karlsruhe, forthcoming

Ontology Learning Layer Cake

$\forall x, y\,(sufferFrom(x, y) \rightarrow ill(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Terms

Terms are at the basis of the ontology learning process

- Terms express more or less complex semantic units
- But what is a term?

Huge Selection of Top Brand Computer Terminals Available for Immediate Delivery

Because Vecmar carries such a large inventory of high-quality computer terminals, including: ADDS terminals, Boundless terminals, DEC terminals, HP terminals, IBM terminals, LINK terminals, NCR terminals and Wyse terminals, your order can often ship same day. Every computer terminal shipped to you is protected with careful packing, including thick boxes. All of our shipping options - including international - are available through major carriers.

- Extracted term candidates (phrases)

  - computer
  - terminal
  - computer terminal
  - ? high-quality computer terminal
  - ? top brand computer terminal
  - ? HP terminal, DEC terminal, …

Term Extraction

Determine most relevant phrases as terms

- Linguistic Methods
  - Rules over linguistically analyzed text
    - Linguistic analysis – Part-of-Speech Tagging, Morphological Analysis, …
    - Extract patterns – *Adjective-Noun, Noun-Noun, Adj-Noun-Noun, …*
    - Ignore *Names* (`DEC`, `HP`, …), *Certain Adjectives* (`quality`, `top`, …), etc.
- Statistical Methods
  - Co-occurrence (collocation) analysis for term extraction within the corpus
  - Comparison of frequencies between domain and general corpora
    - `Computer Terminal` will be specific to the Computer domain
    - `Dining Table` will be less specific to the Computer domain
- Hybrid Methods
  - Linguistic rules to extract term candidates
  - Statistical (pre- or post-) filtering

Linguistic Analysis “Layer Cake”

[[He SUBJ] [booked PRED] [[this] [table HEAD] NP:DOBJ:X1] …] … [[It SUBJ:X1] [was PRED] still available …]	Discourse Analysis
[[He SUBJ] [booked PRED] [[this] [table HEAD] NP:DOBJ] S]	Dependency Struct. (S)
[[the SPEC] [large MOD] [table HEAD] NP]	Dependency Struct. (Phrases)
[[the] [large] [table] NP] [[in] [the] [corner] PP]	Phrase Recognition
[Sommer~schule N] [work~ing V]	Morphological Analysis (“stemming”)
[table N:ARTIFACT] [table N:furniture_01]	PartOfSpeech & Semantic Tagging
[table] [2005-06-01] [John Smith]	Tokenization (incl. Named-Entity Rec.)

Statistical Analysis

Scores used in term extraction:

- MI (Mutual Information) – Cooccurrence Analysis

- TFIDF – Term Weighting

$$tfidf(w) = tf \, . \log(\frac{N}{df(w)})$$

- χ^2 (Chi-square) – Cooccurrence Analysis & Term Weighting

$$X^2 = \sum \frac{(obs - \exp)^2}{\exp}$$

- Other
  - c-value/nc-value (Frantzi & Ananiadou, 1999)
    - Considers length (c-value) and context (nc-value) of terms
  - Domain Relevance & Domain Consensus (Navigli and Velardi, 2004)
    - Considers term distribution within (DC) and between (DR) corpora

TFIDF

most popular weighting schema

(normalized word frequency)

$$tfidf(w) = tf \,.\, \log\left(\frac{N}{df(w)}\right)$$

The word is more important if it appears several times in a target document

The word is more important if it appears in less documents

tf(w)	*term frequency (number of word occurrences in a document)*
df(w)	*document frequency (number of documents containing the word)*
N	*number of all documents*
tfIdf(w)	*relative importance of the word in the document*

Ontology Learning Layer Cake

$\forall x, y\,(sufferFrom\,(x, y) \rightarrow ill\,(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

(Multilingual) Synonyms

- Next step in ontology learning is to identify terms that share (some) semantics, i.e., potentially refer to the same concept

- Synonyms (Within Languages)
  - ‘100% synonyms’ don’t exist – only term pairs with *similar* meanings
  - Examples from http://thesaurus.com
    - `terminal – video display – input device`
    - `graphics terminal - video display unit - screen`

- Translations (Between Languages)
  - ‘100% translations’ don’t exist - only multilingual term pairs with *similar* meanings
  - Examples from http://dict.leo.org
    - `input device` (English) – `Eingabegerät` (German)
    - Back to English: `input device, input unit, signal conditioning device`
    - `video display unit` (English) – `Videosichtgerät` (German)

Extraction of Synonyms

Term Classification and Clustering

- Classification
  - Classifying terms to existing class systems, e.g., by extending WordNet (with SynSets corresponding to classes)
- Clustering
  - Clusters according to similar distributions, e.g., by measuring co-occurrence between terms

Extraction of Translations

Multilingual Term Classification and Clustering - see e.g. Grefenstette, 1998

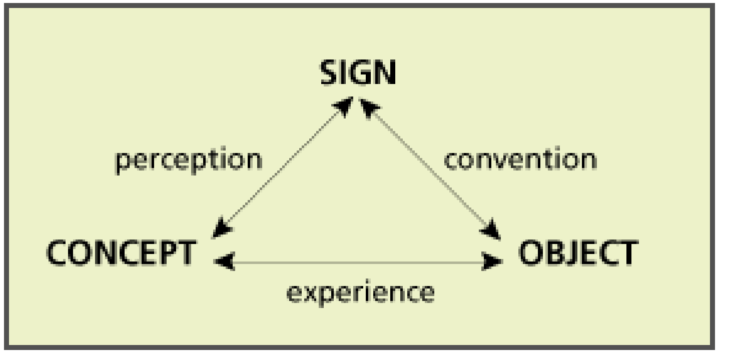
- Similar as with *monolingual* terms, but depending on translated contexts (i.e., document collections):
  - Parallel Corpora: Pairs of translated documents
  - Comparable Corpora: Pairs of documents in different languages on the same topic
- In both cases ‘need to cross the language barrier’
  - Parallel Corpora: Term alignment according to document structure (layout, linguistic, semantic)
  - Comparable Corpora: Term alignment according to similar contexts, e.g. by translating context words (dictionary lookup)

Ontology Learning Layer Cake

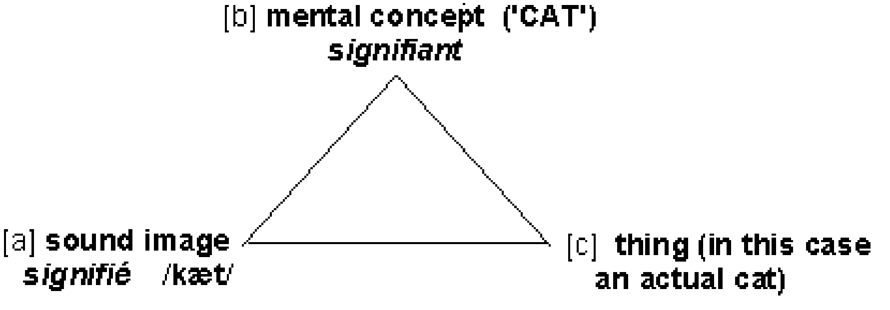
$\forall x, y\,(sufferFrom(x, y) \rightarrow ill(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

The Semiotic Triangle

Ogden & Richards, 1923



- based on Structural Linguistics studies (de Saussure, 1916)



- adopted in Knowledge Representation (e.g. Sowa, 1984)

Concepts: *Intension, Extension, Lexicon*

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

- Intension
  - (in)formal definition of the set of objects that this concept describes
    - *a disease is an impairment of health or a condition of abnormal functioning*

- Extension
  - a set of objects (instances) that the definition of this concept describes
    - *influenza, cancer, heart disease, …*

Discussion: what is an instance? - '*heart disease*' or '*my uncle's heart disease*'

- Lexical Realizations
  - the term itself and its multilingual synonyms
    - *disease, illness, Krankheit, maladie, …*

Discussion: synonyms vs. instances – '*disease*', '*heart disease*', '*cancer*', …

Concepts: *Intension*

Extraction of a Definition for a Concept from Text

- Informal Definition
  - e.g., a gloss for the concept as used in WordNet
  - *OntoLearn* (Navigli and Velardi, 2004; Velardi et al., 2005) uses natural language generation to compositionally build up a WordNet gloss for automatically extracted concepts
    - 'Integration Strategy' : "*strategy for the integration of …*"

- Formal Definition
  - e.g., a logical form that defines all formal constraints on class membership
  - Inductive Logic Programming, Formal Concept Analysis, …

Concepts: *Extension*

Extraction of Instances for a Concept from Text

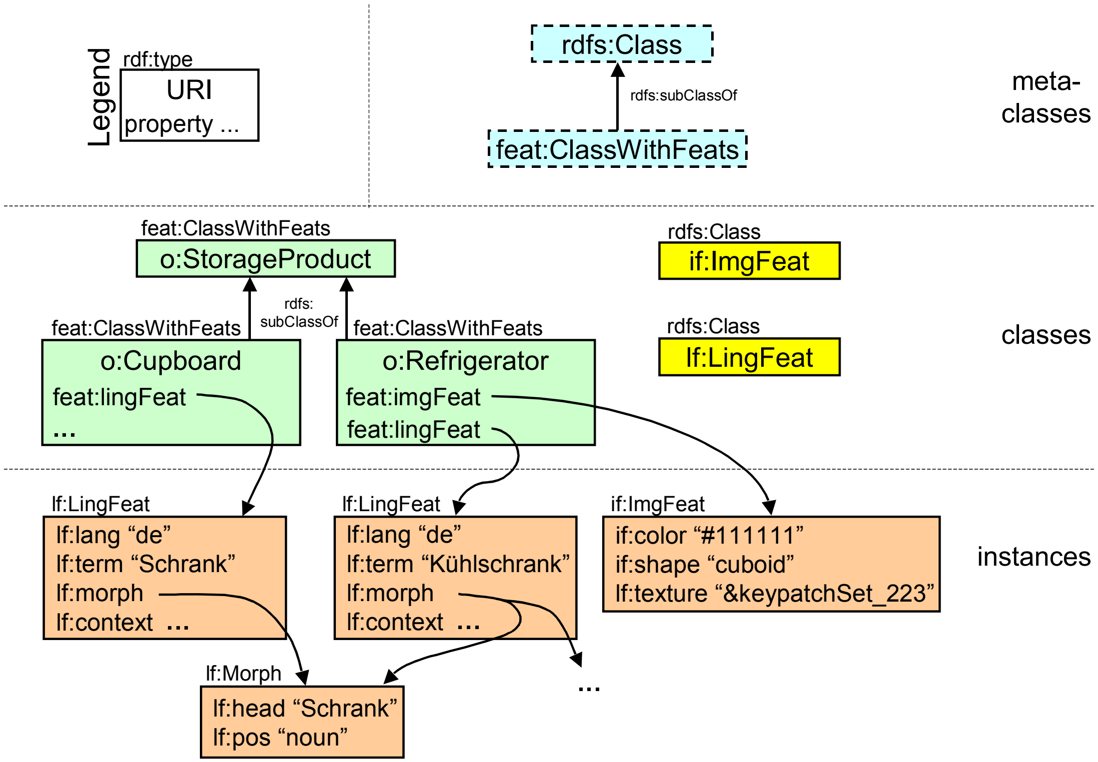
- Commonly referred to as Ontology Population
- Relates to Knowledge Markup (Semantic Metadata)
- Uses Named-Entity Recognition and Information Extraction

- Instances can be:
  - Names for objects, e.g.
    - *Person, Organization, Country, City, …*
  - Event instances (with participant and property instances), e.g.
    - *Football Match (with Teams, Players, Officials, ...)*
    - *Disease (with Patient-Name, Symptoms, Date, …)*

Concepts: *Lexicon*

Extraction of Synonyms and Translations for a Concept from Text

- (Multilingual) Term Extraction – see previous slides
- Representation of Lexical Information in Ontologies (Buitelaar et al., 2005)



Ontology Learning Layer Cake

$\forall x, y\,(sufferFrom(x, y) \rightarrow ill(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Taxonomy Extraction - Overview

- **Lexico-syntactic patterns**
- Distributional Similarity & Clustering
- Linguistic Approaches
- Document-subsumption
- Taxonomy Extension/Refinement
- Combination Opportunities

Hearst Patterns [Hearst 1992]

- Examples for hyponymy patterns:
  - *Vehicles **such as** cars, trucks and bikes*
  - ***Such** fruits **as** oranges, nectarines or apples*
  - *Swimming, running **and other** activities*
  - *Publications, **especially** papers and books*
  - *A seabass **is** a fish.*

Hearst Patterns [Hearst 1992]

- Examples for hyponymy patterns:
  - *NP **such as** NP, NP, ... and NP*
  - ***Such** NP **as** NP, NP, ... or NP*
  - *NP, NP, ... **and other** NP*
  - *NP, **especially** NP, NP ,... and NP*
  - *NP **is** a NP.*
  - *...*
- Principle idea: match these patterns in texts to retrieve isa-relations
- Precision wrt. Wordnet: 55,46% (66/119)

Extensions of Hearst's approach

- Using Hearst Patterns for Anaphora Resolution
  - Poesio et al. 02 / Markert et al. 03
- Additional Patterns
  - [Iwanska et al. 00]
- Using Questions
  - [Sundblad 02]
- Application to collateral texts
  - [Ahmad et al. 03]
- Matching patterns on the Web
  - KnowItAll [Etzioni et al. 04-05], PANKOW [Cimiano et al. 04-05]
- Improving Accuracy (LSA) & Coverage (Conjunctions)
  - [Cederberg and Widdows 03]
- Learning Patterns
  - Snowball [Agichtein et al. 00], [Downey et al. 04], [Ravichandran and Hovy 02], [Snow et al. 04])

Taxonomy Extraction - Overview

- Lexico-syntactic patterns
- **Distributional Similarity & Clustering**
- Linguistic Approaches
- Document-subsumption
- Taxonomy Extension / Refinement
- Combination Opportunities

Distributional Hypothesis & Vector Space Model

- Harris, 1986
  - *„Words are (semantically) similar to the extent to which they share similar words“*
- Firth, 1957
  - *„You shall know a word by the company it keeps“*

- Idea: collect context information and represent it as a vector:

	book_obj	rent_obj	drive_obj	ride_obj	join_obj
apartment	X	X			
car	X	X	X		
motor-bike	X	X	X	X	
excursion	X				X
trip	X				X

- compute similarity among vectors wrt. a measure

Context Features

- **Four-grams** [Schuetze 93]
- **Word-windows** [Grefenstette 92]
- **Predicate-Argument relations** (*every man loves a woman*)
  Modifier Relations (*fast car, the hood of the car*)
  - [Grefenstette 92, Cimiano 04b, Gasperin et al. 03]
- **Appositions** (*Ferrari, the fastest car in the world*)
  - [Caraballo 99]
- **Coordination** (*ladies and gentlemen*)
  - [Caraballo 99, Dorow and Widdows 03]

Using Syntactic Surface Dependencies

*Mopti is the **biggest** city **along** the Niger with one of the most **vibrant** ports and a large bustling market. Mopti has a **traditional** ambience that other towns **seem** to have lost. It is also the center **of** the **local** tourist industry and **suffers from** hard-sell overload. The **nearby** junction towns of Gao and San **offer nice** views **over** the Niger's delta.*

```
city: biggest(1)
ambience: traditional(1)
center: of_tourist_industry(1)
junction town: nearby(1)
market: bustling(1)
port: vibrant(1)
overload:suffer_from(1)
tourist industry: center_of(1), local(1)
town: seem_subj(1)
view: nice(1), offer_obj(1)
```

Clustering Concept Hierarchies from Text

- **Similarity-based**
- Set-theoretical and Probabilistic
- Soft clustering

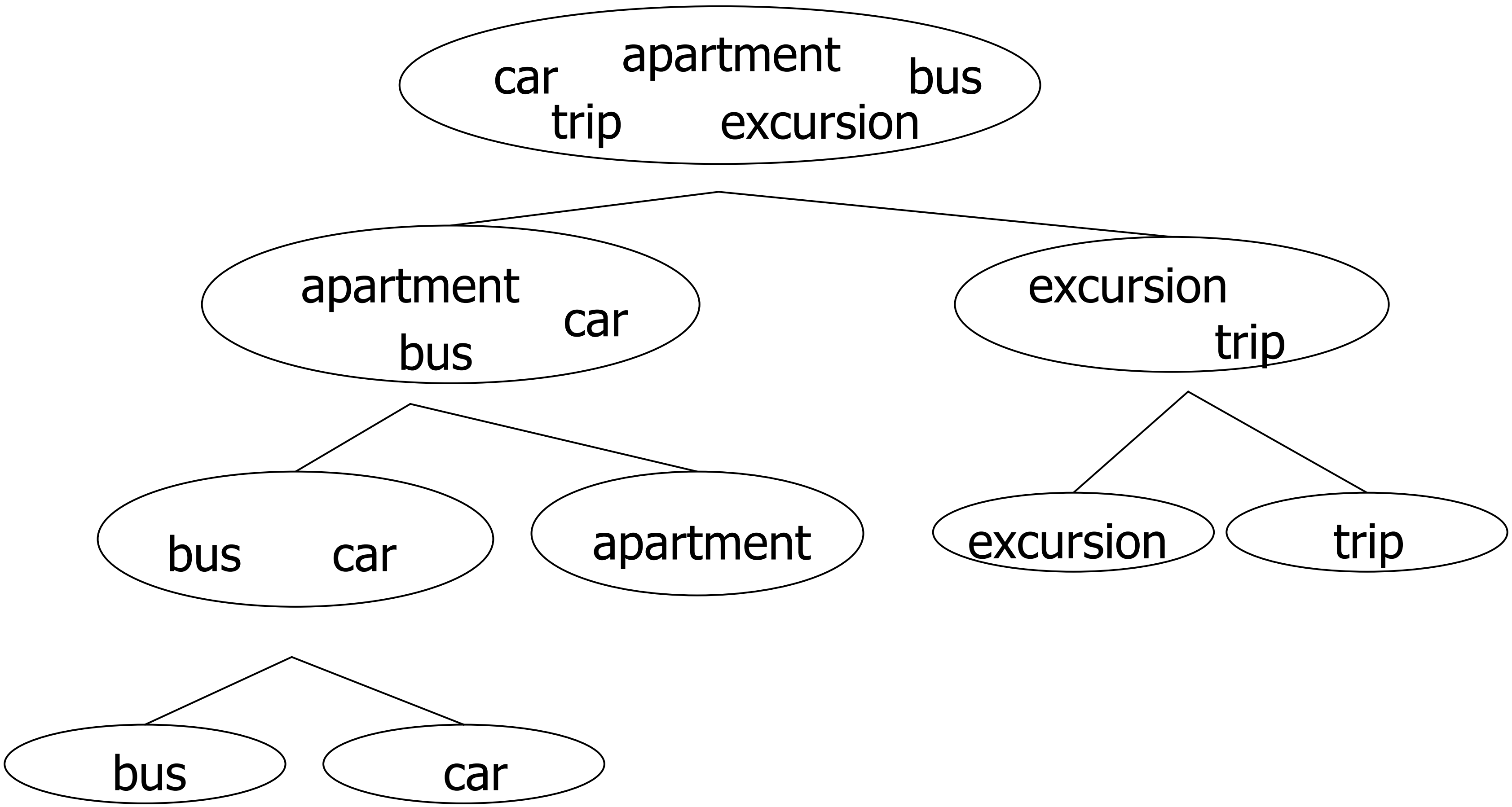
Similarity-based Clustering

- Similarity Measures:
  - Binary (Jaccard, Dine)
  - Geometric (Cosine, Euclidean/Manhattan distance)
  - Information-theoretic (Relative Entropy, Mutual Information)
  - (…)

- Linkage Strategies:
  - Complete linkage
  - Average linkage
  - Single linkage
  - (…)

- Methods:
  - Hierarchical agglomerative clustering
  - Hierarchical top-down clustering, e.g. Bi-Section KMeans
  - (…)

Bi-Section-KMeans



Problem 1: Labeling of Clusters

- Caraballo's Method [1999]:
  - Agglomerative Clustering
  - Labeling Clusters with hypernyms derived from Hearst patterns
  - Removing unlabeled concepts thus compacting the hierarchy

- Evaluation: select 20 nouns with at least 20 hypernyms and present them to human judges with the 3 best hypernyms for each

- Results:
  - Best Hypernym (33% (Majority) / 39% (Any)
  - Any Hypernym (47.5% (Majority) / 60.5% (Any))

Problem 2: Spurious Similarities

- Guided Clustering [Cimiano 2005c]:
  - Integrate a externally derived hypernym oracle into the agglomerative clustering algorithm
  - Two terms are only clustered if they have a common hypernym according to the oracle
  - Label the cluster with the common hypernym
    - ⇒ Demonstrably better hierarchies
    - ⇒ Labels for the cluster

⇒ Reuse techniques from Clustering with constraints!

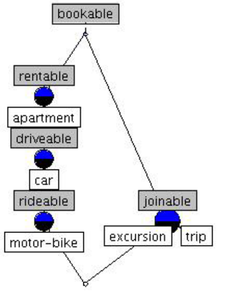
Clustering Concept Hierarchies

- Similarity-based
- **Set Theoretical & Probabilistic**
- Soft clustering

Set Theoretical & Probabilistic Clustering

- Set theoretical
  - Formal Concept Analysis [Ganter and Wille 1999]

	bookable	rentable	drivable	ridable	joinable
apartment	X	X			
car	X	X	X		
motor-bike	X	X	X	X	
excursion	X				X
trip	X				X



- COBWEB [Fisher 87]
  - probabilistic representation of features
  - incremental clustering
  - hill-climbing search

Clustering – Comparison [Cimiano 04]

	F-Measure	Time Complexity	Understandability
FCA	43.81/41.02%	$O(2^n)$	Good
Agglomerative Clustering	36.78/33.35% 36.55/32.92% 38.57/32.15%	$O(n^2 \log(n))$ $O(n^2)$ $O(n^2)$	Fair
Divisive Clustering	36.42/32.77%	$O(n^2)$	Weak-Fair

Clustering Concept Hierarchies from Text

- Similarity-based
- Set-theoretical & Probabilistic
- **Soft clustering**

What About Multiple Word Meanings?

- **bank**: financial institute or natural object?
  - At least two clusters!
- So we need soft clustering algorithms:
  - Clustering By Committee (CBC) [Lin et al. 2002]
  - Gaussian Mixtures (EM)
  - **PoBOC** (Pole-Based Overlapping Clustering)
  - FCA
  - (...)
- Challenge: recognize multiple word meanings!

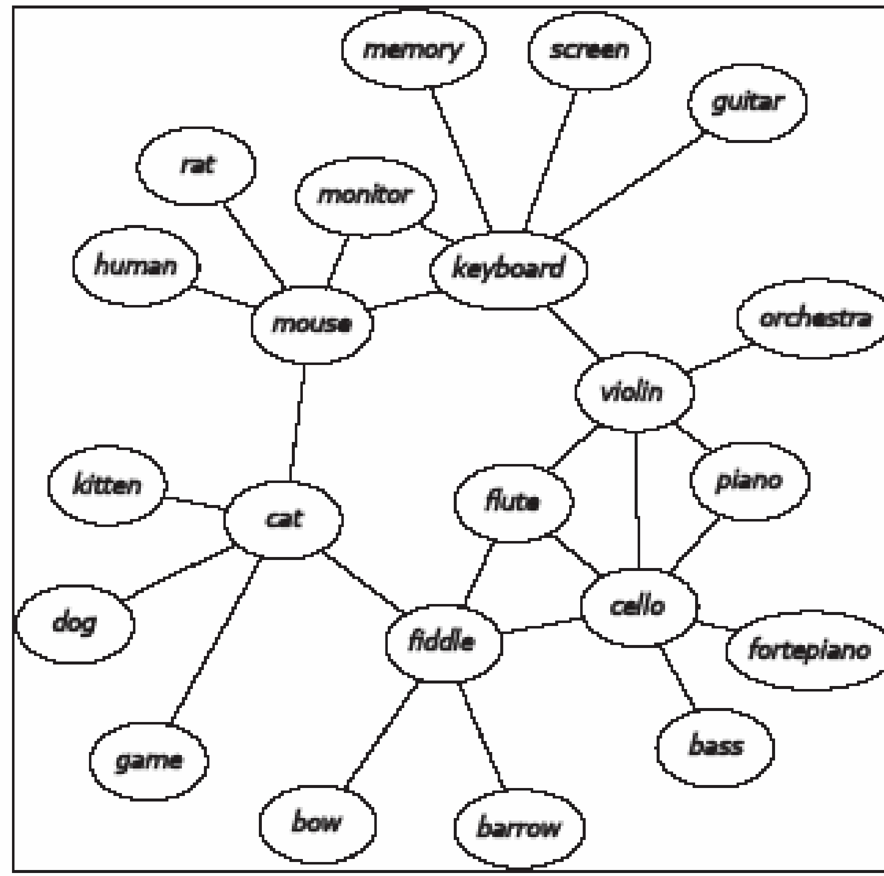
Approach by [Widdows and Dorow 2002]

Use coordination patterns:

- keyboards and pianos.
- A mouse and a cat.

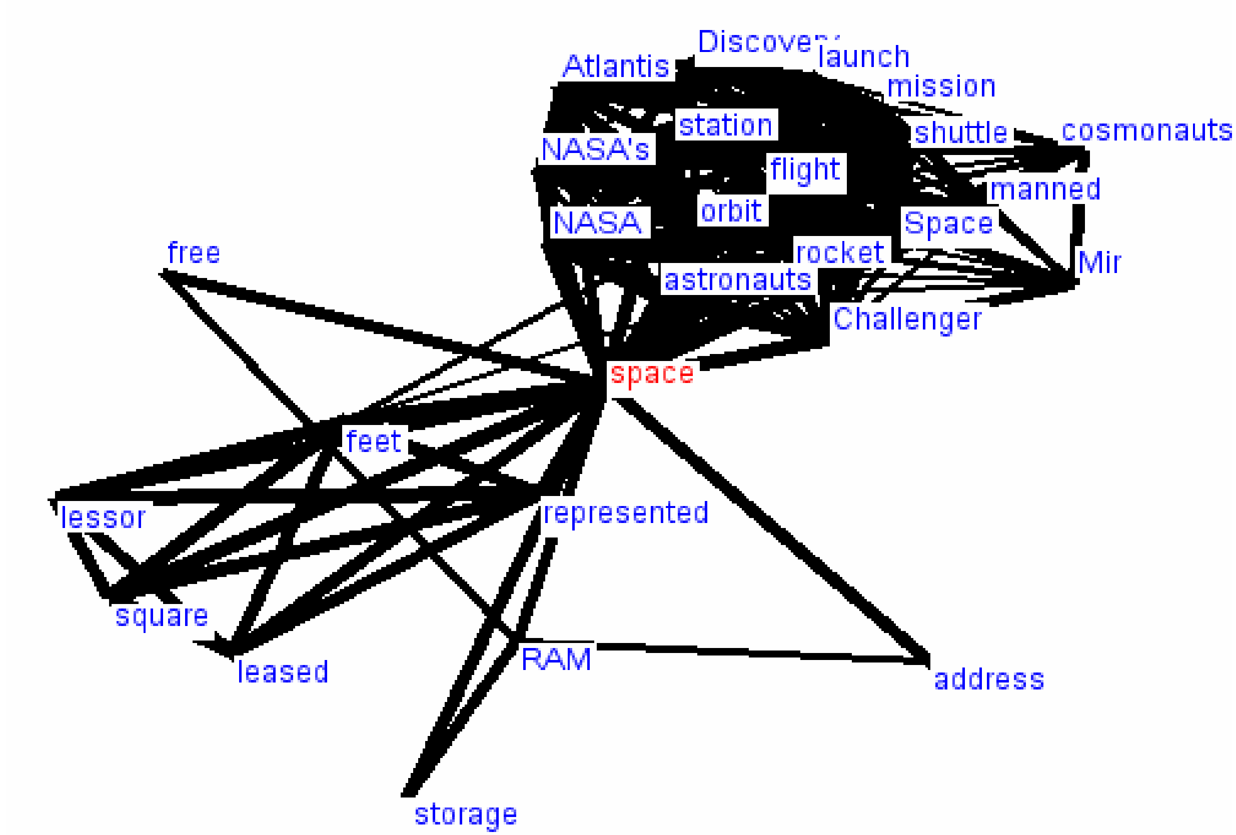
Apply LSA/LSI to reduce dimension of co-occurence matrix.

Calculate similarity as the cosine between the angle of the corresponding vectors



Use of Collocations „Deutscher Wortschatz"-Project

Collocations: *„A occurs together with B more than expected by chance"*



Taxonomy Extraction - Overview

- Lexico-syntactic patterns
- Distributional Similarity & Clustering
- **Linguistic Approaches**
- Document subsumption
- Taxonomy Extension / Refinement
- Combination Opportunities

Linguistic Approaches

- Modifiers:
  - Modifiers (adjectives/nouns) typically restrict or narrow down the meaning of the modified noun, i.e.
  - e.g. *isa(international credit card, credit card)*
  - Yields a very accurate heuristic for learning taxonomic relations, e.g. OntoLearn [Velardi&Navigli], OntoLT [Buitelaar et al., 2004], TextToOnto [Cimiano et al.], [Sanchez et al., 2005]

- Compositional interpretation of compounds [OntoLearn]
  - e.g. *long-term debt*
    - Disambiguate *long-term* and *debt* with respect to WordNet
    - Generate a gloss out of the glosses of the respective synsets:

      long-term debt := *„a kind of debt, the state of owing something (especially money), relating to or extending over a relatively long time"*

Taxonomy Extraction - Overview

- Lexico-syntactic patterns
- Distributional Similarity & Clustering
- Linguistic Approaches
- **Document subsumption**
- Taxonomy Extension / Refinement
- Combination Opportunities

Approach by [Sanderson and Croft]

- A term t1 subsumes a term t2, i.e. is-a(t2,t1) if t1 appears in all the documents in which t2 appears [Sanderson and Croft 1999]

- Probabilistic definition [Fotzo 04]:

is-a(t2,t1) iff P(t1|t2) > t

$$P(x \mid y) = \frac{n(x,y)}{n(y)}$$

where n(x, y) is the number of documents in which x and y co - occur and

n(y) is the number of documents in which y occurs

Taxonomy Extraction - Overview

- Lexico-syntactic patterns
- Distributional Similarity & Clustering
- Linguistic Approaches
- Document subsumption
- **Taxonomy Extension/Refinement**
- Combination Opportunities

Taxonomy Extension/Refinement

Approach	Technique	Accuracy	Learning Accuracy
Widdows 03	LSA (Wordspace)	10%	?
Alfonseca et al. 02	Signatures	17.39%	38%
Maedche, Pekar & Staab 02	Tree-Ascending+ kNN	15.74%	39.46%
Witschel 05	Decision Trees	11-14%	40-60%

Conclusions:

- difficult problem
- approaches not comparable (datasets, measures, ontologies, number of concepts,...)

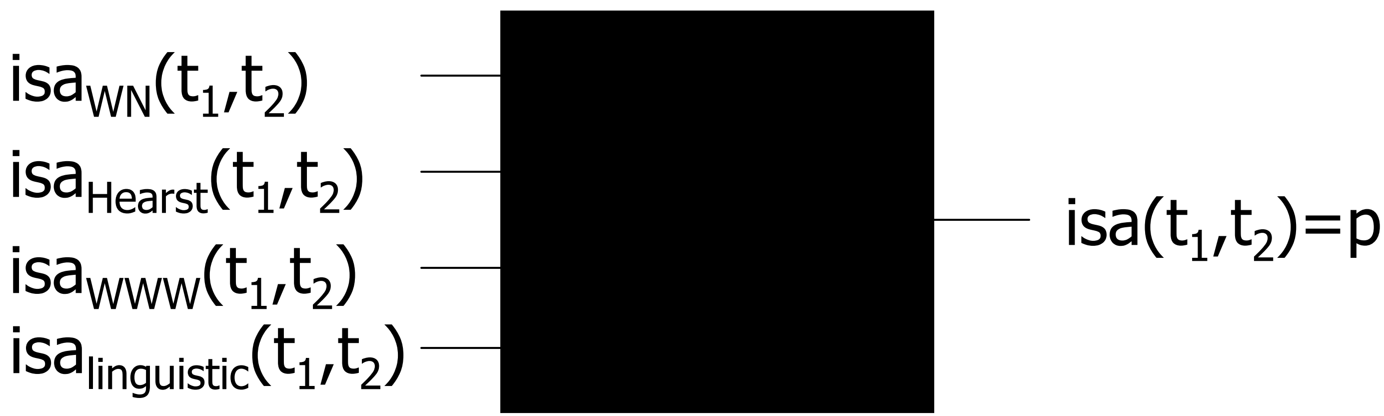
Taxonomy Extraction - Overview

- Lexico-syntactic patterns
- Distributional Similarity & Clustering
- Linguistic Approaches
- Document subsumption
- Taxonomy Extension / Refinement
- **Combination Opportunities**

Initial Blueprints for Combination

- [Caraballo 99]
  - Label tree produced with hierarchical agglomerative clustering using lexico-syntactic patterns
- [Cimiano 05b/c]
  - Guided Clustering
    - Integrate a hypernym oracle with agglomerative clustering
  - Classification-based approach
    - use features derived from several learning paradigms
- [Cederberg & Widdows 03]
  - Increase accuracy and coverage of lexico-syntactic patterns by using LSA and coordination patterns

Classification-based approach



<u>Idea:</u> Use as input features derived by applying different techniques, resources, etc. and find optimal combination in a supervised manner!

Ontology Learning Layer Cake

Example	Layer
$\forall x, y\,(sufferFrom(x, y) \rightarrow ill(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Specific Relations / Attributes

- Part-of [Charniak et al. 98]
  - X consists of Y
- Qualia [Yamada et al. 04, Cimiano & Wenderoth 05]
  - Formal: such X as Y
  - Purpose: X is used for Y
  - Agentive: a ADV Xed Y
- Causation [Girju 02]
  - X leads to Y
- Attributes [Poesio and Almuhareb 05]
  - the X of Y

General Relations: Exploiting Linguistic Structure

- **OntoLT**: *SubjToClass_PredToSlot_DObjToRange* Heuristic
  - Maps a linguistic subject to a class, its predicate to a corresponding slot for this class and the direct object to the range of the slot

- **TextToOnto**: Acquisition of Subcategorization Frames, e.g.
  - love(man,woman)
  - love(kid,mother)
  - love(kid,grandfather)

  love(person,person)

- Problem related to acquisition of *subcategorization frames* and *selectional restrictions* [Resnik 97, Ribas 95, Clark and Weir 02] in Natural Language Processing

Which Relations are Actually the Same?

- Clustering of verbs semantically according to their alternation behavior [Schulte im Walde 00]
- Use EM algorithm
- Examples:
  - {advise, teach, instruct}
  - {fly, move, roll}
  - {start, finish, stop, begin}
  - {fight, play}
  - {meet, play}
  - {need, like, want , desire}

Finding the Right Level of Abstraction

- [Ciramita et al. 05]
  - Genia Corpus. + Genia Ontology
  - Verb-based relations
    - X activates B
- Use X^2 to decide to generalize or not (significance level)
- Results:
  - 83.3% of relations correct according to human evaluation
  - 53.1% correctly generalized

Ontology Learning Layer Cake

$\forall x, y\,(sufferFrom\,(x, y) \rightarrow ill\,(x))$	Rules & Axioms
cure(dom:DOCTOR,range:DISEASE)	Relations
is_a(DOCTOR,PERSON)	Taxonomy
DISEASE:=<Int,Ext,Lex>	Concepts
{disease, illness, Krankheit}	(Multilingual) Synonyms
disease, illness, hospital	Terms

Axioms

- DIRT (Discovery of Inference Rules from Text: Lin et al. 2001)
  - calculate significant collocations on dependency paths
  - Examples: „X solves Y“
    - Y is solved by X, X resolves Y, X finds a solution to Y, X tries to solve Y, Y deals with X, Y is resolved by X, X addresses Y, X seeks a solution to Y, X do something about Y, ...
- AEON [Völker et al. 2005]:
  - Rigidity, Identity, Unity, Dependence
- [Haase and Völker 2005]
  - Disjointness Axioms on the basis of coordination:
    - i.e. disjoint(man,woman)

Part IV

Ontology Evaluation

based on the „Ontology Evaluation" SEKT Report
by Janez Brank, Marko Grobelnik, Dunja Mladenić (2005)

Towards Ontology Evaluation

- A key factor which makes a particular discipline scientific is the ability to **evaluate and compare the ideas** within the area.
  - …the same holds also for Semantic Web research area when dealing with abstractions in the form of ontologies.

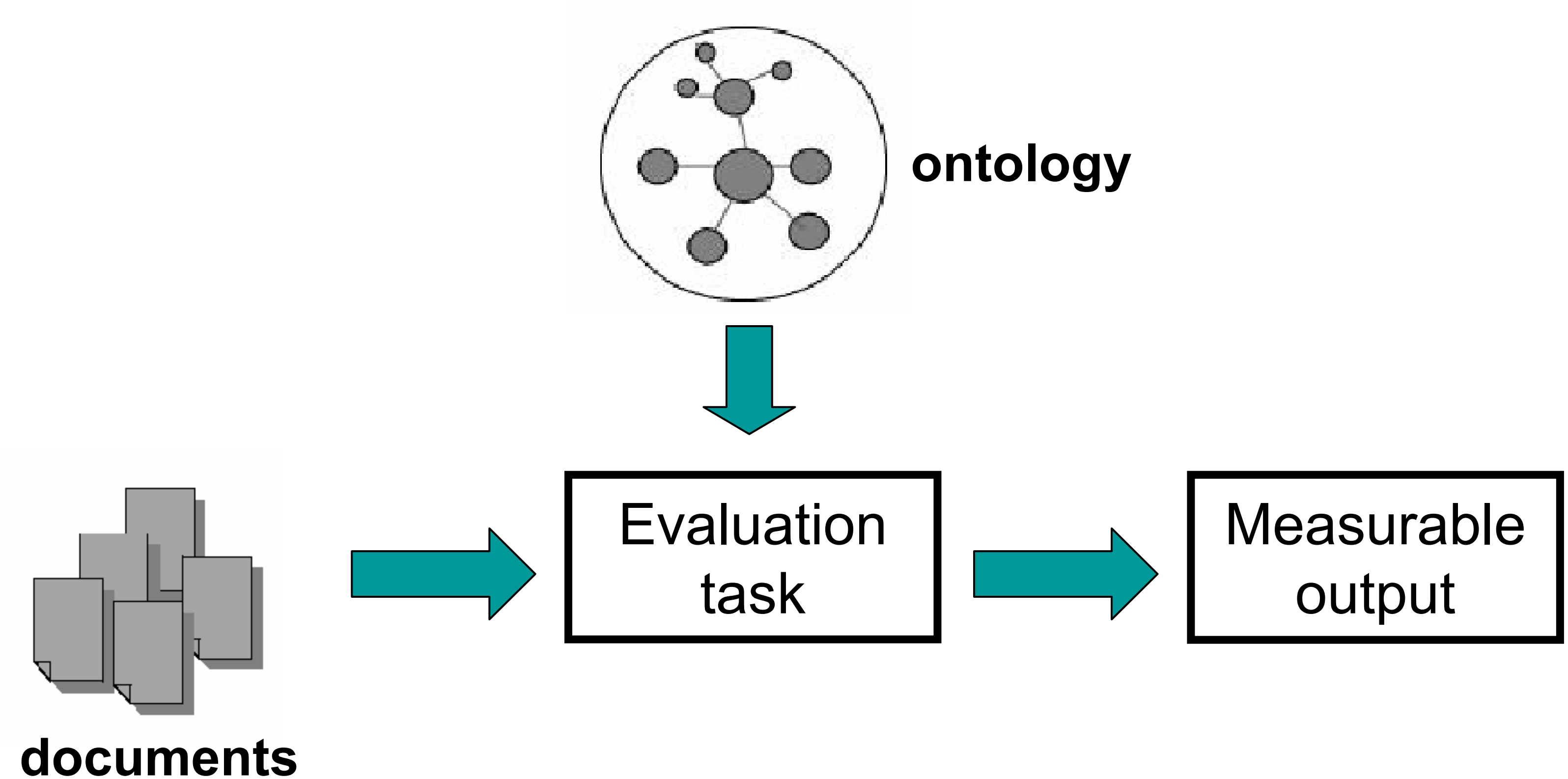
- Ontologies are fundamental data structures for conceptualizing knowledge which are in most practical cases **non-uniquely expressible**
  - …as a consequence, we can build many different ontologies conceptualizing the same body of knowledge and should be able to say which of them **serve better their purpose**.

Why Evaluate Ontologies?

- Ontology evaluation could be important in several contexts (e.g.):
  - A user may be wondering which ontology in a given library is most **suitable for given requirements**;
  - …or **how good an ontology has been produced** by some ontology construction effort (either manual or automated);
  - …or evaluation can be a **component in automated ontology learning approaches** for guiding the exploration within a search space.

Typical Scenario When Evaluating Ontologies

(…but not necessarily the only possible)



based
stand

ba

involving
don

evaluation is done by humans who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc

	A… ches t… ati…			
Levels	**Golden standard**	**Application-based**	**Data-driven**	**Assessment by humans**
Lexical, vocabulary, data	x	x	x	x
Hierarchy, taxonomy	x	x	x	x
Other semantic relations	x	x	x	x
Context, application		x		x
Syntactic	x			x
Structure, architecture, design				x
Philosophical				x

Common Approaches to Ontology Evaluation

- Evaluation approaches fall into one of the following categories:
  - comparing the ontology to a "**golden standard**" (which may itself be an ontology; e.g. Maedche and Staab, 2002)
  - using the **ontology in an application** and evaluating the results (e.g. Porzel and Malaka, 2004)
  - involving **comparisons with a source of data** about the domain that is to be covered by the ontology (e.g. Brewster *et al*., 2004)
  - evaluation is **done by humans** who try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc. (e.g. Lozano-Tello and Gómez-Pérez, 2004)

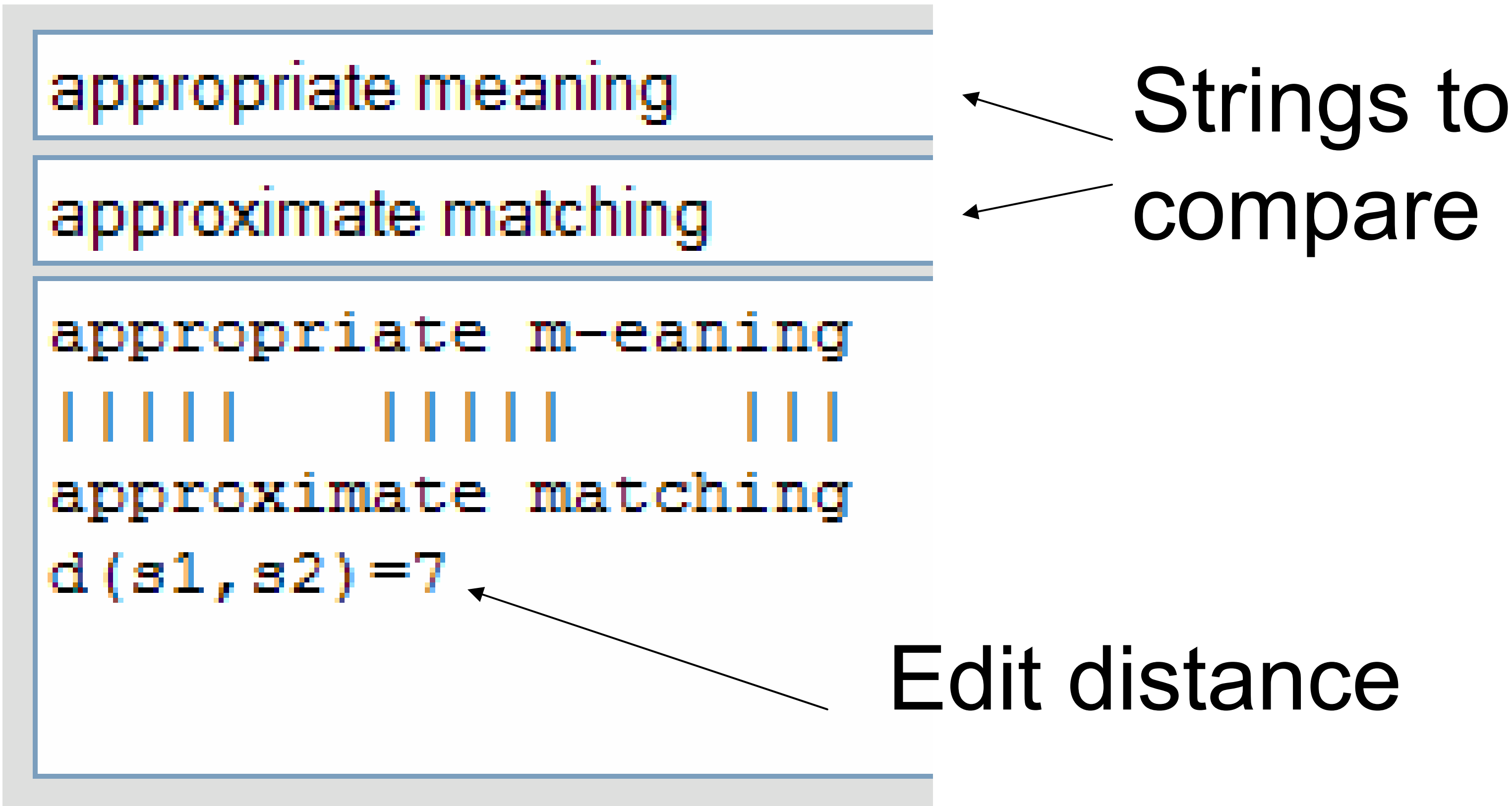
Lexical, Vocabulary, Data

	Approaches to evaluation			
Levels	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

String Distances for Ontology Evaluation

- Maedche and Staab (2002)
  - Similarity between two strings is measured based on the Levenshtein edit distance, normalized to produce scores in the range [0, 1]
    - background knowledge (such as abbreviations) could be used
  - **A *string matching* measure between two sets of strings is then defined by taking each string of the first set, finding its similarity to the most similar string in the second set, and averaging this over all strings of the first set.**
  - This is used for taking the set of all strings used as concept identifiers in the ontology being evaluated, and compare it to a "golden standard" set

Edit Distance Example



Precision/Recall for Ont. Evaluation

- Lexical content of an ontology can also be evaluated using the concepts of precision and recall (as known in Information Retrieval)
  - Precision would be the percentage of terms (strings used as concept identifiers) that also appear in the golden standard, relative to the total number of terms
  - Recall is the percentage of the golden standard terms that also appear as concept identifiers in the ontology, relative to the total number of golden standard terms

Glosses/Patterns for Ontology Evaluation

- (Velardi *et al.* 2005) approach extracts relevant domain-specific concepts, and finds definitions for them (using web-search and WordNet entries) and connects some of the concepts by is-a relations:
  - Part of their evaluation approach is to generate natural-language glosses for multiple-word terms
  - The glosses are of the form:

    "*x y* = a kind of *y*, ⟨definition of *y*⟩, related to the *x*, ⟨definition of *x*⟩"
  - A gloss like this would then be shown to human domain experts, who would evaluate it to see if the word sense disambiguation algorithm selected the correct definitions of ***x*** and ***y***.

Hierarchy, Taxonomy

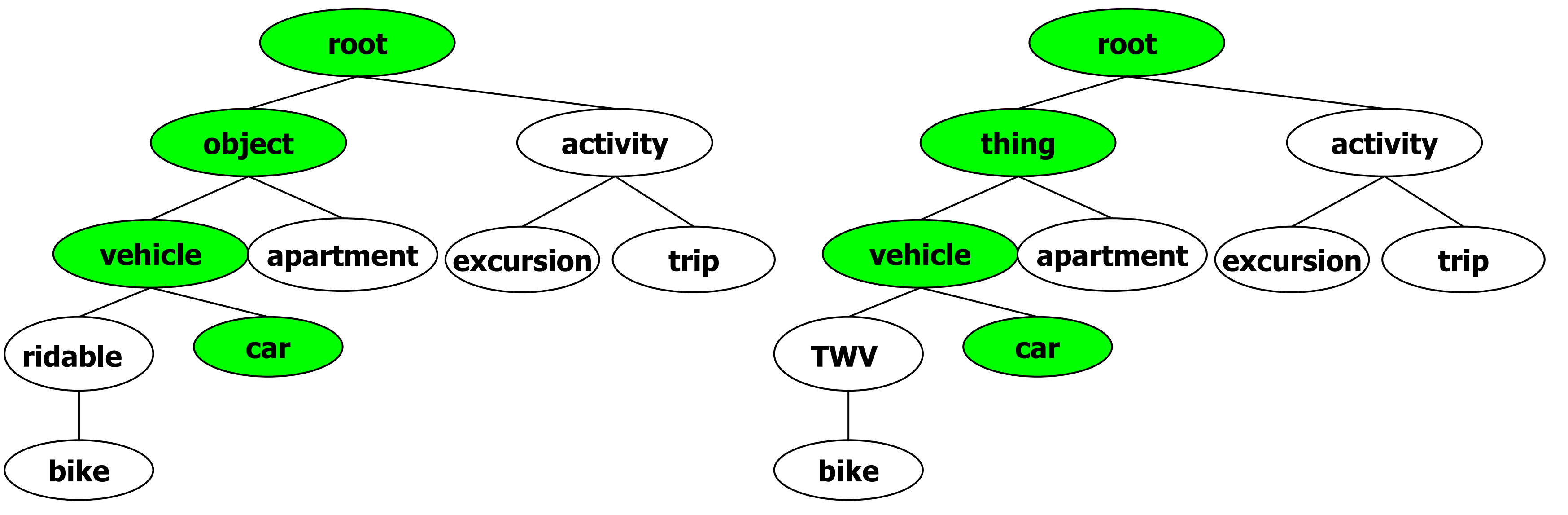
	Approaches to evaluation			
Levels	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

Semantic Cotopy [Maedche and Staab, 2002]

- *Semantic cotopy* of a term *c* in a given hierarchy is the set of all its super- and sub-concepts
  - Given two hierarchies $\leq_{C1}$ and $\leq_{C2}$
    - The overlap of the semantic cotopy of c1 in $\leq_{C1}$ as well as the semantic cotopy of c2 in $\leq_{C2}$ can be used as a measure of how similar both concepts c1 and c2 are.
    - An average of this may then be computed over all the terms occurring in the two hierarchies; **this is a measure of similarity between $\leq_{C1}$ and $\leq_{C2}$.**

Def. & Example for Semantic Cotopy

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



=> $TO(car, O_1, O_2) = 3/4$

Other Semantic Relations

	Approaches to evaluation			
Levels	**Golden standard**	**Application-based**	**Data-driven**	**Assessment by humans**
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

Structural Fit [Brewster et al., 2004]

- Data-driven approach to evaluate the degree of structural fit between an ontology and a doc. corpus:
  - EM clustering is performed on corpus of documents
  - Each concept *c* of the ontology is represented by a set of terms
  - The clusters (in the form of probabilistic models) representing topics can be used to measure, how well a concept *c* form ontology fits that topic
  - Concepts associated with the same topic should be closely related in the ontology (via is-a and possibly other relations).
    - …this would indicate that the structure of the ontology is reasonably well aligned with the hidden structure of topics in the domain-specific corpus of documents

Context, Application

	Approaches to evaluation			
Levels	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

How Context is Used for Evaluation

- Ontology could be a part of a larger collection of ontologies that may reference one another
  - e.g. one ontology may use a class or concept declared in another ontology
  - Possible scenarios are on the web or within some institutional library of ontologies.

- This context can be used for evaluation of an ontology in various ways
  - The Swoogle portal [Ding et al., 2004] and OntoKhoj portal of [Patel et al., 2003] redefine the well known PageRank algorithm according to the link structure between semantic-web documents
    - …context is provided through external link structure (how other people link our concepts)
  - [Supekar, 2005] proposes semantic search based on context provided by humans

Swoogle
Ding *et al.* (2004)

- Swoogle search engine uses cross-references between semantic-web documents to define a graph and then compute a score for each ontology in a manner analogous to PageRank

- …the resulting "ontology rank" is used to rank query results

Philosophical

	Approaches to evaluation			
Levels	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

Guarino and Welty (2002) (1/2)

- They point out several philosophical notions (essentiality, rigidity, unity, etc.) that can be used to better understand the nature of conceptualizations
- Example:
  - a property is said to be *essential* to an entity if it necessarily holds for that entity.
  - …a property that is essential for all entities having this property is called *rigid*
    - (e.g. “being a person”: there is no entity that could be a person but isn’t; everything that is a person is necessarily always a person)
  - …a property that cannot be essential to an entity is called *anti-rigid*
    - (e.g. “being a student”: any entity that is a student could also not be a student)

Guarino and Welty (2002) (2/2)

- This approach could be used for detecting of, e.g., various other kinds of misuse of the is-a relationship
- A downside of this approach is that it requires manual intervention by a trained human expert
- Völker *et al.* (2005) recently proposed an approach to aid in the automatic assignment of these metadata tags

Multiple Criteria Approaches

	Approaches to evaluation			
Levels	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, data	X	X	X	X
Hierarchy, taxonomy	X	X	X	X
Other semantic relations	X	X	X	X
Context, application		X		X
Syntactic	X			X
Structure, architecture, design				X
Philosophical				X

How Multiple Criteria are Used

- Ontologies are evaluated using several decision criteria or attributes:
  - …for each criterion, the ontology is evaluated and given a numerical score
  - …additionally a weight is assigned to each criterion, and an overall score for the ontology is then computed as a weighted sum of its per-criterion scores
- Next two slides include two sets of possible criteria

Examples of Multiple Criteria Burton-Jones *et al.* (2004)

- **lawfulness** (i.e. frequency of syntactical errors)
- **richness** (how much of the formal language is actually used in ontology)
- **interpretability** (do the terms used in the ontology also appear in WordNet)
- **consistency** (how many concepts in the ontology are inconsistent)
- **clarity** (do the terms used in the ontology have many senses in WordNet)
- **comprehensiveness** (number of concepts in the ontology, relative to the average for the entire library of ontologies)
- **accuracy** (percentage of false statements in the ontology)
- **relevance** (number of statements that involve syntactic features marked as useful or acceptable to the user/agent)
- **authority** (how many other ontologies use concepts from this ontology),
- **history** (how many accesses to this ontology have been made, relative to other ontologies in the library/repository)

Examples of Multiple Criteria Fox *et al.* (1998)

- **functional completeness** (does the ontology contain enough information for the application at hand)
- **generality** (is it general enough to be shared by multiple users, departments, etc.)
- **efficiency** (does the ontology support efficient reasoning)
- **perspicuity** (is it understandable to the users)
- **precision/granularity** (does it support multiple levels of abstraction/detail)
- **minimality** (does it contain only as many concepts as necessary)

Summary of Ontology Evaluation

- We presented Ontology Evaluation through:
  - …different approaches
  - …on different levels
- The main aim of doing evaluation is to be able to find better conceptualization for the same corpus of knowledge
  - …evaluation measures are used to guide such a search

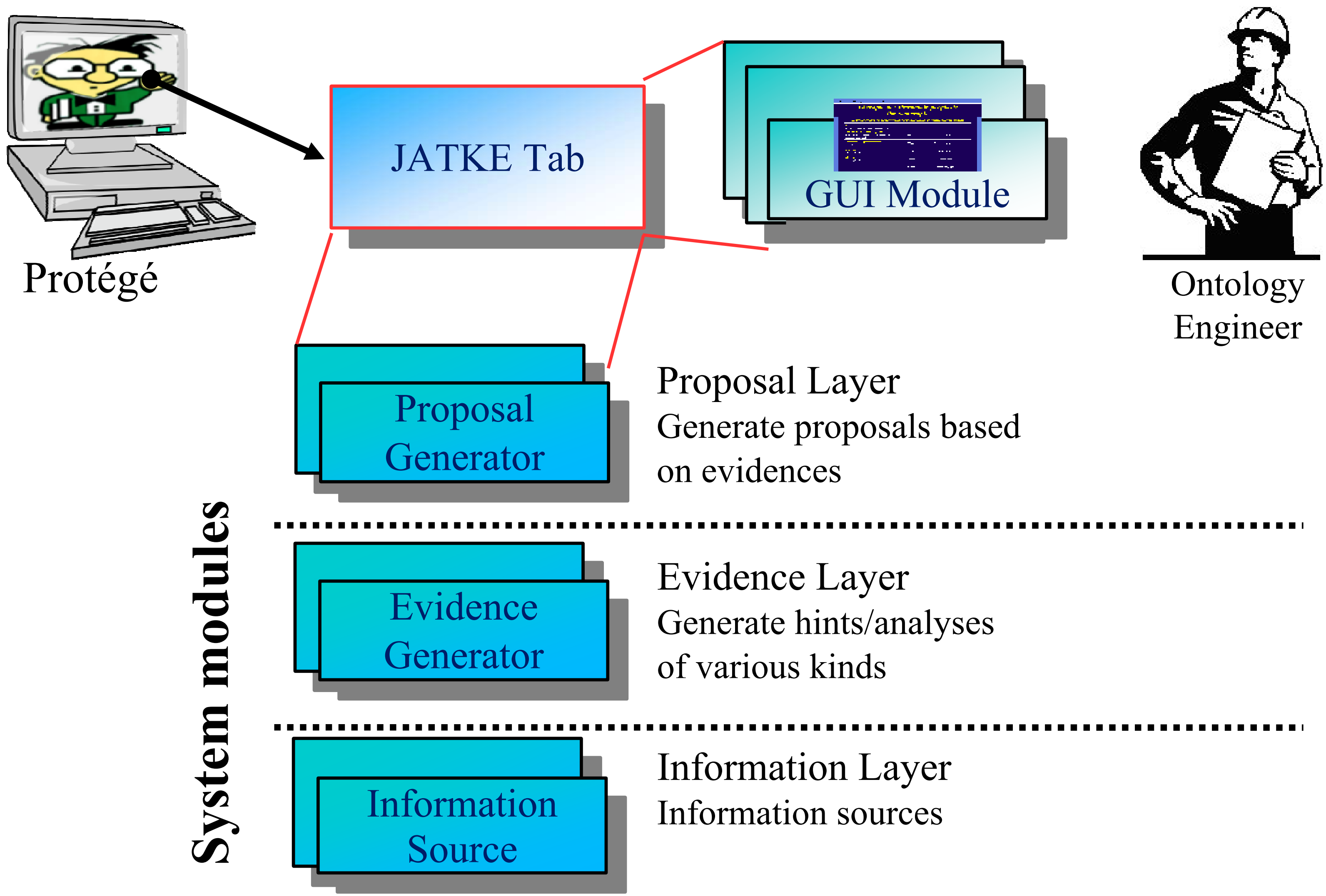
Part V

Tools for Ontology Learning from Text

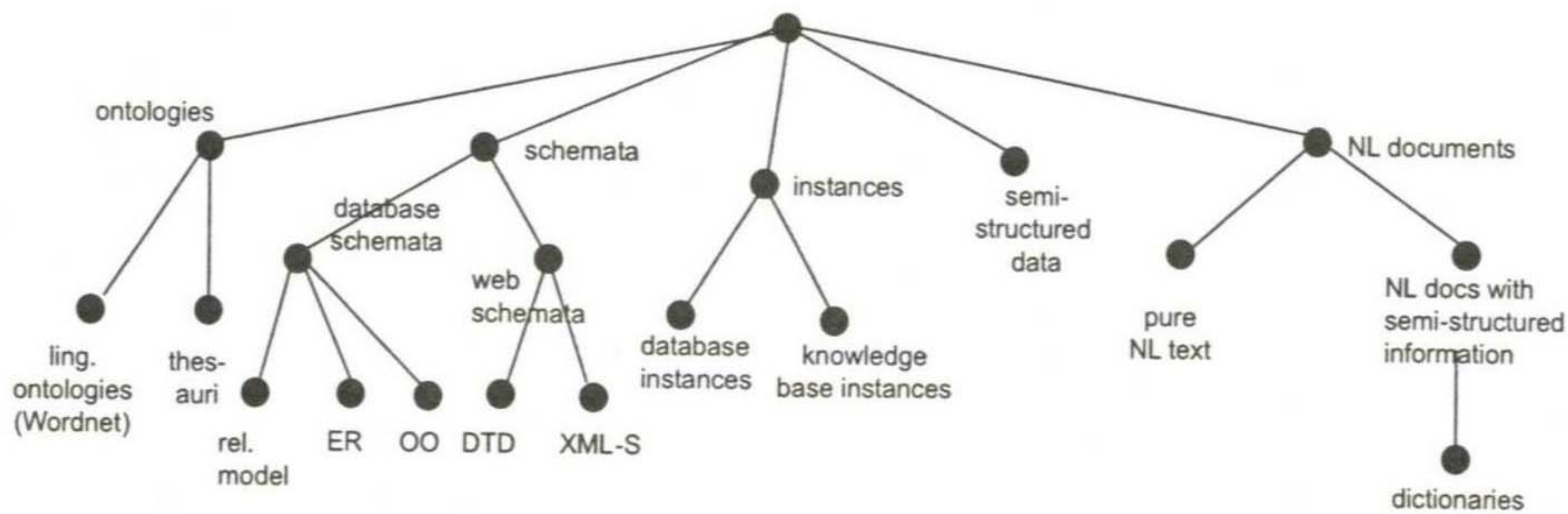
JATKE: A Framework for Ontology Learning (DFKI *Knowledge Management Dept*.)

- Allows combination (via plugins) of various methods for ontology learning, e.g.
  - Statistics-based
  - Structure-based
  - NLP-based
- Methods generate evidences from various information sources (ontologies, documents, user feedback, …) which are used to propose ontology changes to the user
- Availability: open source (Java, Protégé Plugin)
- Link: *http://jatke.opendfki.de*

JATKE: Module Structure

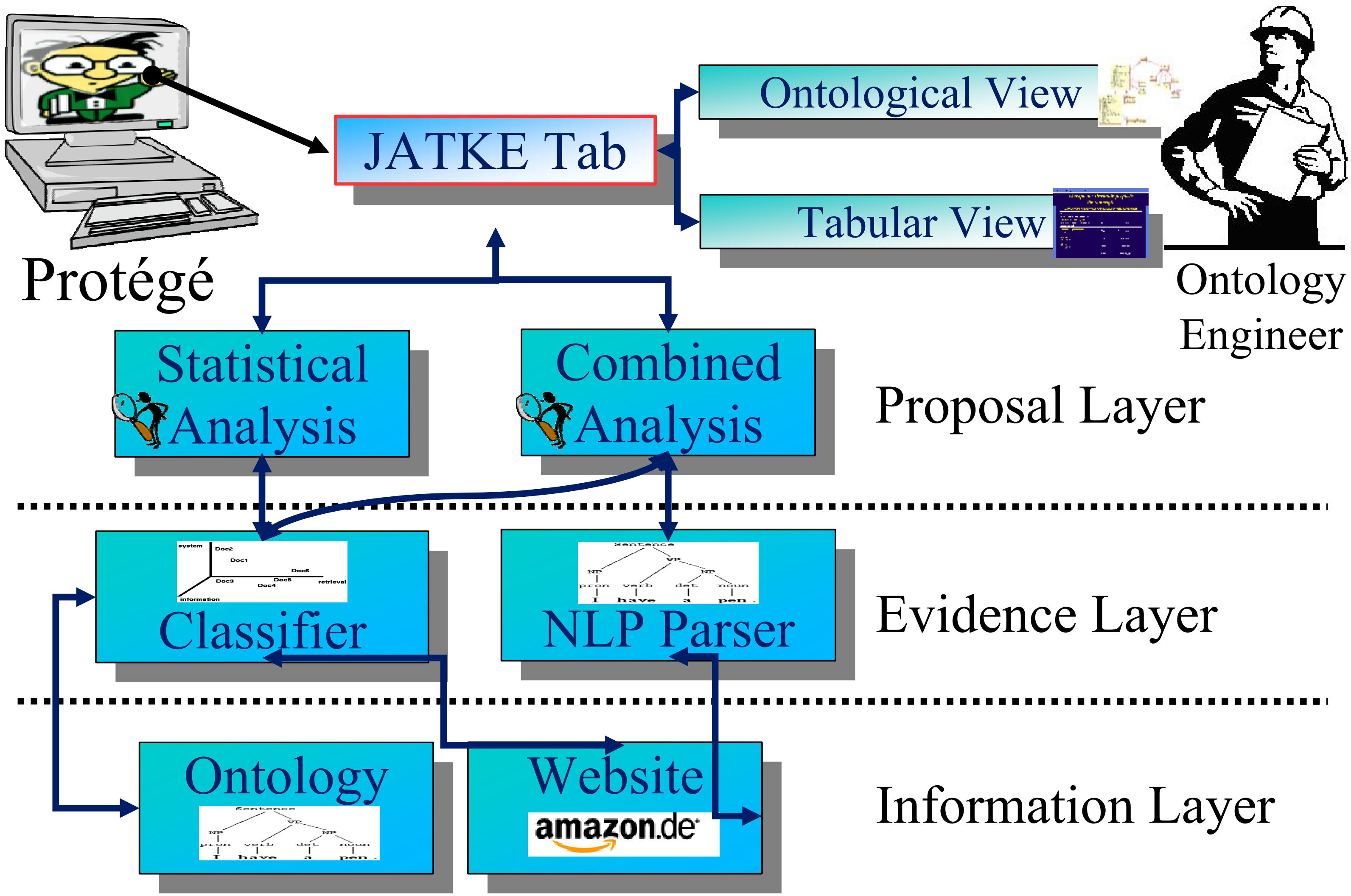


Information Layer

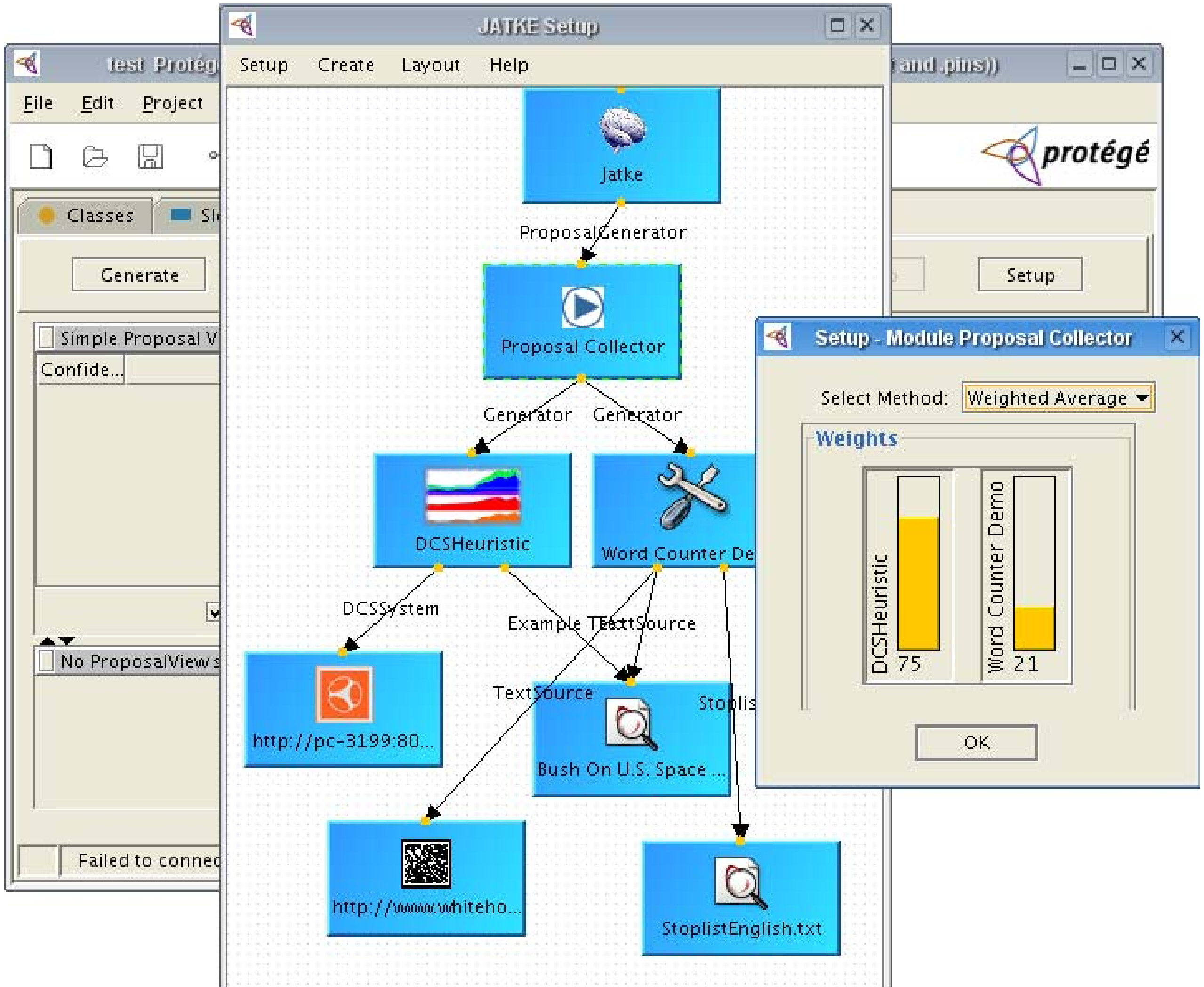


Taxonomy of Relevant Data for Ontology Learning
(from A. Maedche "Ontology Learning for the Semantic Web", PHD Thesis)

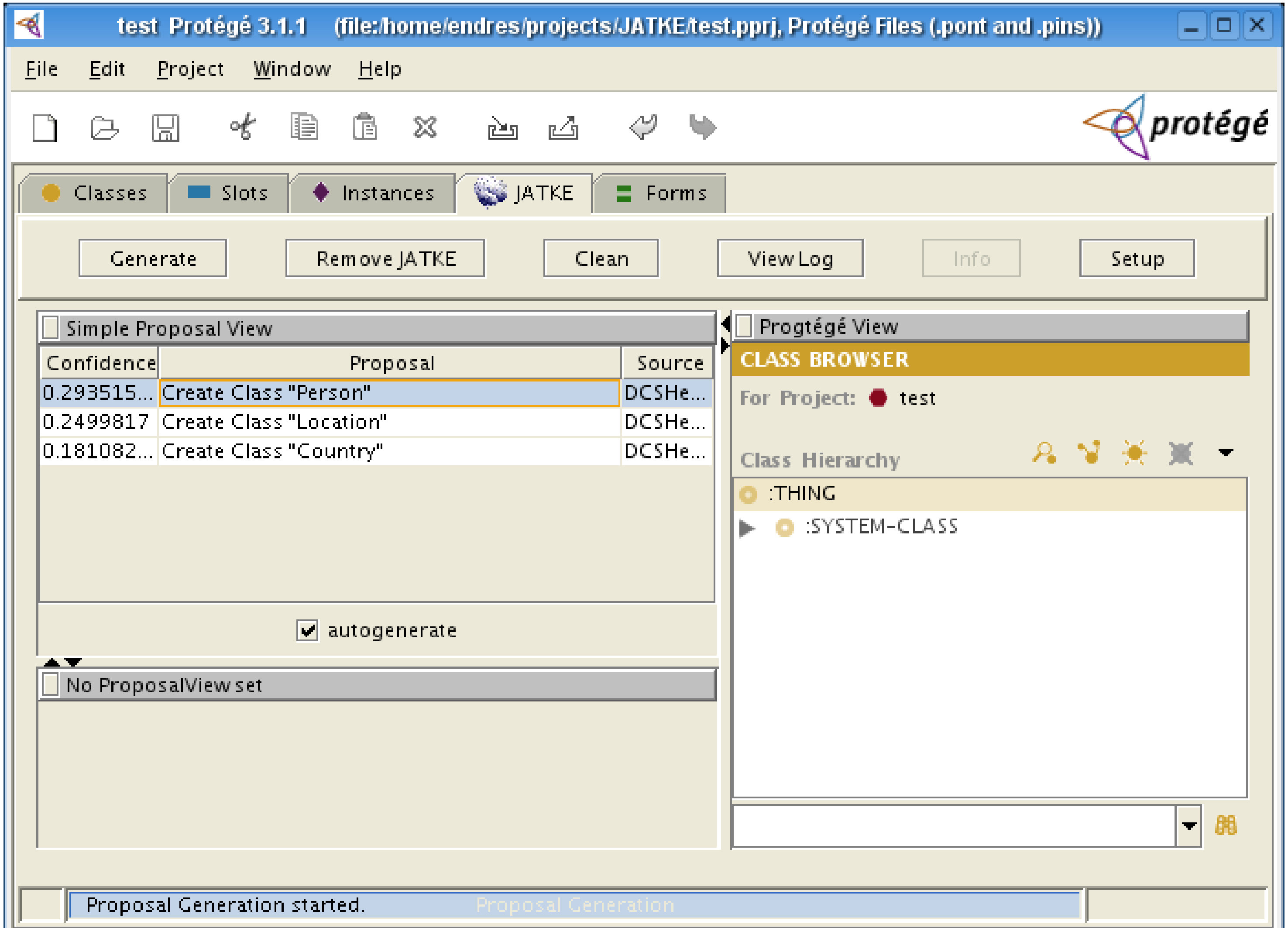
JATKE: Configuration Example



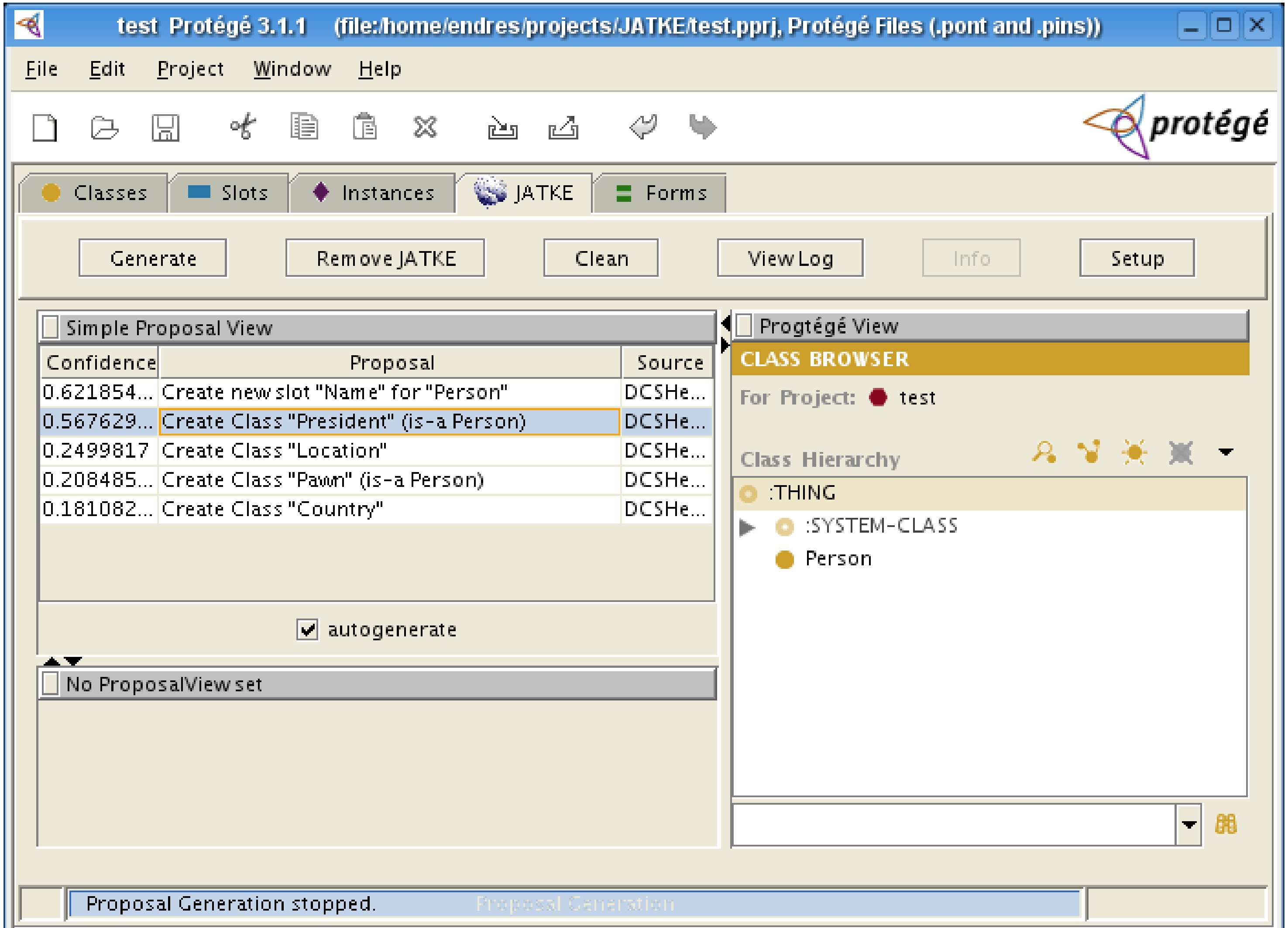
JATKE: Screenshots



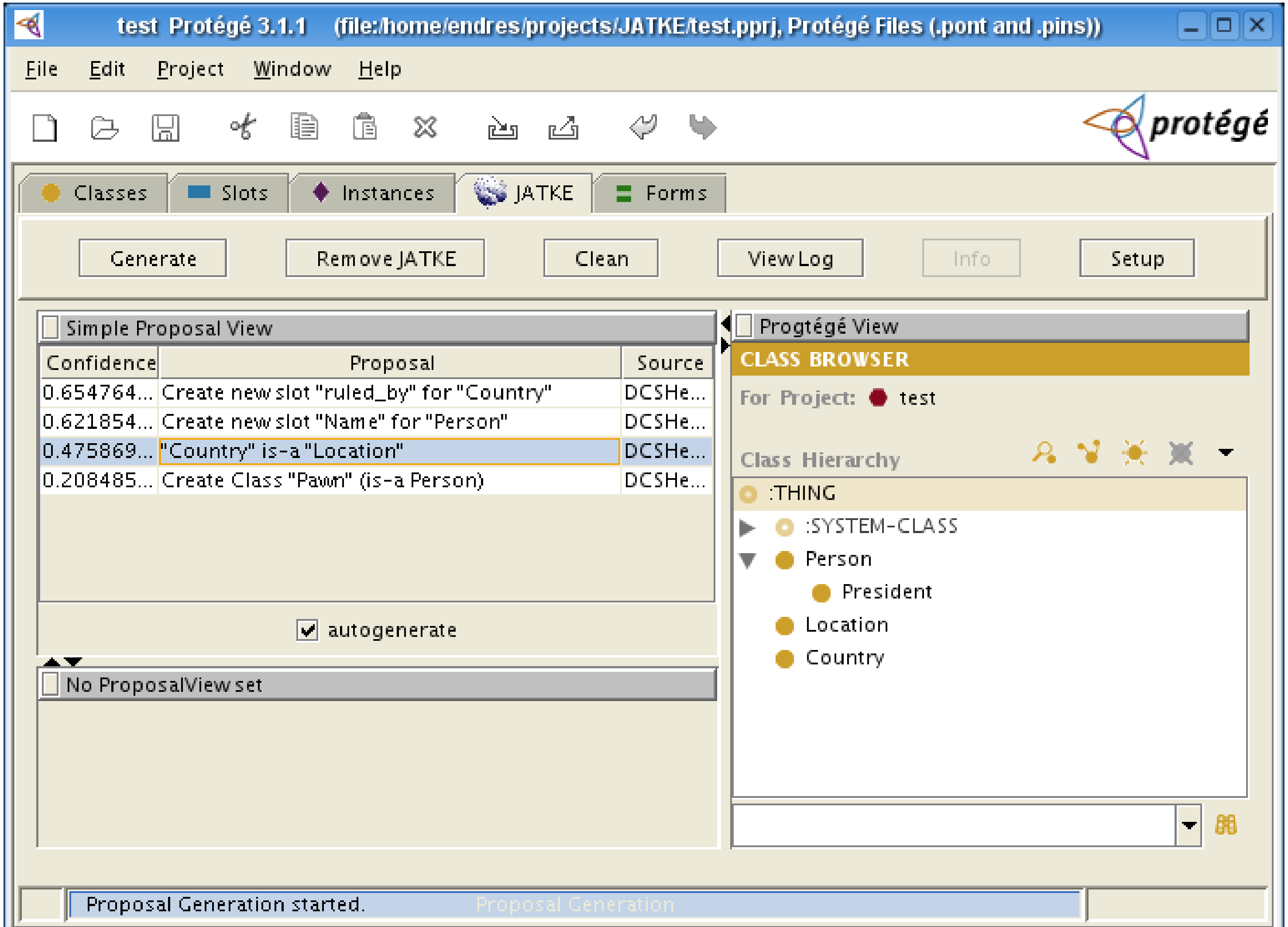
JATKE in Action



JATKE in Action



JATKE in Action



TextToOnto (AIFB, University of Karlsruhe)

- Main features:
  - Taxonomy induction using conceptual clustering (FCA)
  - Taxonomy induction using a combination of techniques
  - Learning subcategorization frames for relation learning
  - Learning Relations by mining association rules

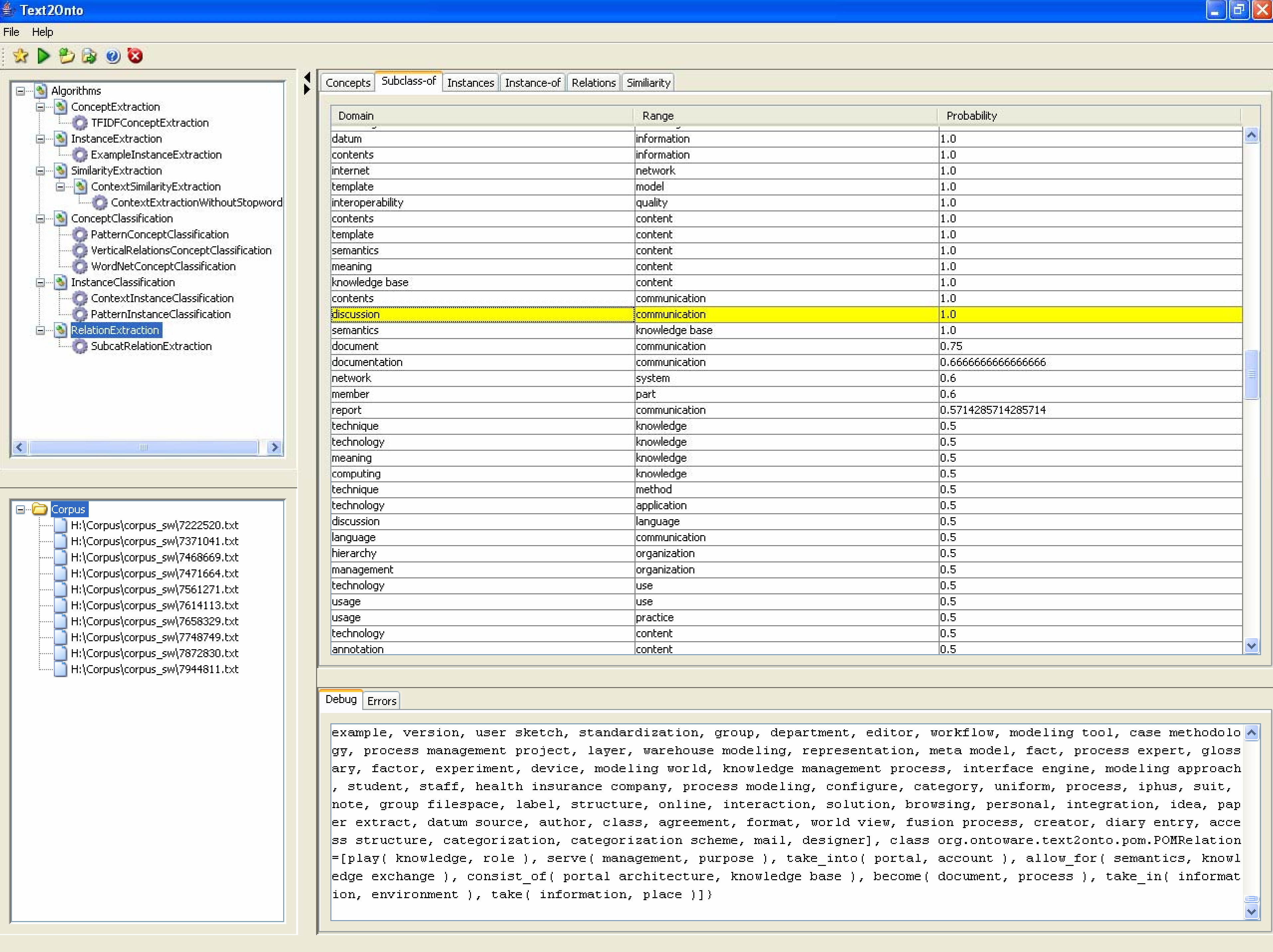
- Other Features:
  - Corpus Management
  - Ontology Editor
  - KAON as ontology repository

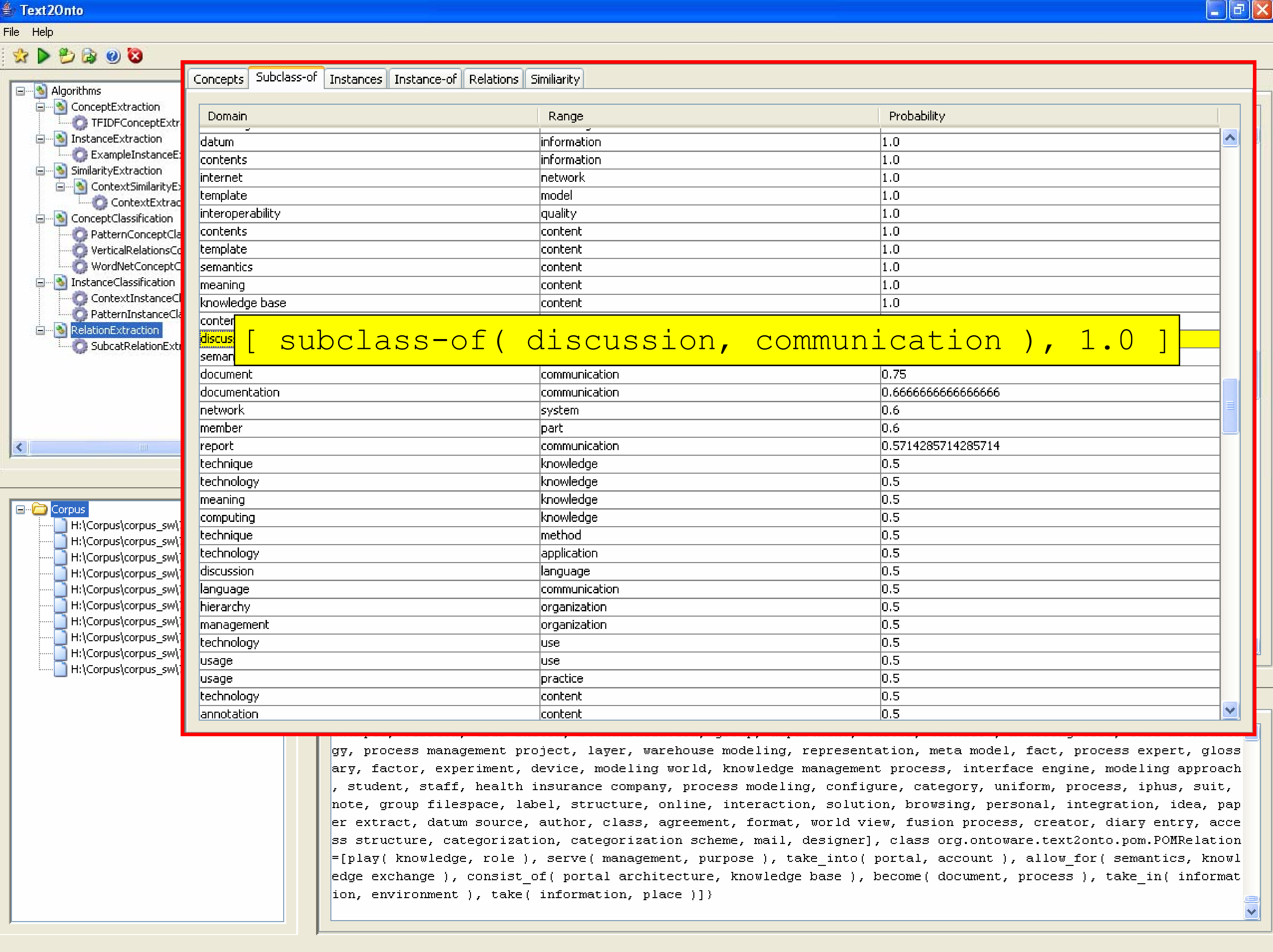
- Availability: open source (Java)
- Link: *http://sourceforge.net/projects/texttoonto*

Text2Onto (AIFB, University of Karlsruhe)

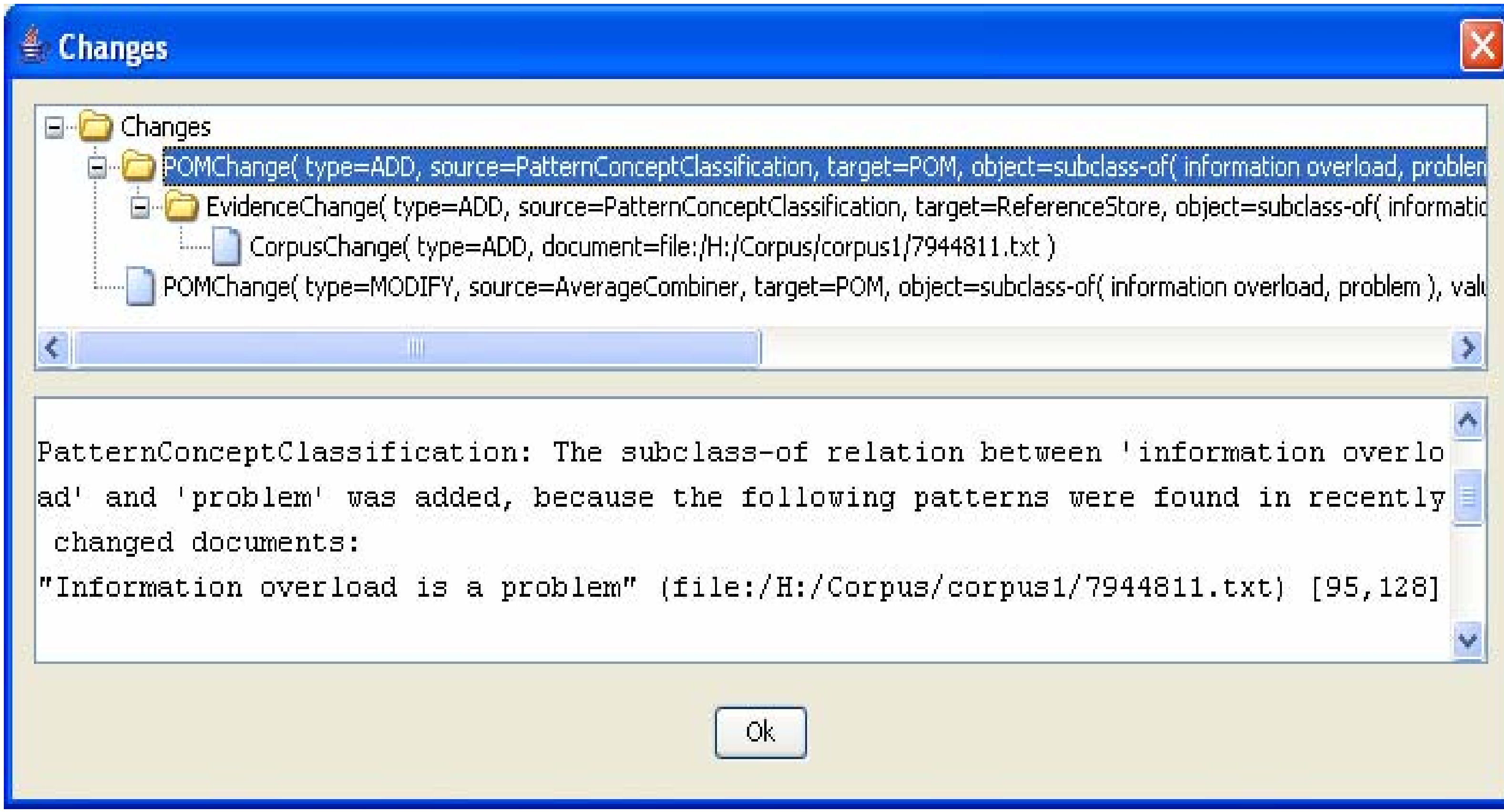
- Main features:
  - Track ontology changes with respect to corpus changes
  - Efficiency by incremental learning
  - Explanation component
  - Learn primitives independent of a specific KR language
  - Confidences for better user interaction
  - allows for easy:
    - combination of algorithms
    - execution of algorithms
    - writing of new algorithms

- Availability: open source (Java)
- Link: http://ontoware.org/projects/text2onto/


Text2Onto
File Help
Algorithms
ConceptExtraction
TFIDFConceptExtraction
InstanceExtraction
ExampleInstanceExtraction
SimilarityExtraction
ContextSimilarityExtraction
ContextExtractionWithoutStopword
ConceptClassification
PatternConceptClassification
VerticalRelationsConceptClassification
WordNetConceptClassification
InstanceClassification
ContextInstanceClassification
PatternInstanceClassification
RelationExtraction
SubcatRelationExtraction
Corpus
H:\Corpus\corpus_sw\7222520.txt
H:\Corpus\corpus_sw\7371041.txt
H:\Corpus\corpus_sw\7468669.txt
H:\Corpus\corpus_sw\7471664.txt
H:\Corpus\corpus_sw\7561271.txt
H:\Corpus\corpus_sw\7614113.txt
H:\Corpus\corpus_sw\7658329.txt
H:\Corpus\corpus_sw\7748749.txt
H:\Corpus\corpus_sw\7872830.txt
H:\Corpus\corpus_sw\7944811.txt
Concepts Subclass-of Instances Instance-of Relations Similiarity
Domain Range Probability
datum information 1.0
contents information 1.0
internet network 1.0
template model 1.0
interoperability quality 1.0
contents content 1.0
template content 1.0
semantics content 1.0
meaning content 1.0
knowledge base content 1.0
contents communication 1.0
discussion communication 1.0
semantics knowledge base 1.0
document communication 0.75
documentation communication 0.6666666666666666
network system 0.6
member part 0.6
report communication 0.5714285714285714
technique knowledge 0.5
technology knowledge 0.5
meaning knowledge 0.5
computing knowledge 0.5
technique method 0.5
technology application 0.5
discussion language 0.5
language communication 0.5
hierarchy organization 0.5
management organization 0.5
technology use 0.5
usage use 0.5
usage practice 0.5
technology content 0.5
annotation content 0.5
Debug Errors
example, version, user sketch, standardization, group, department, editor, workflow, modeling tool, case methodology, process management project, layer, warehouse modeling, representation, meta model, fact, process expert, glossary, factor, experiment, device, modeling world, knowledge management process, interface engine, modeling approach, student, staff, health insurance company, process modeling, configure, category, uniform, process, iphus, suit, note, group filespace, label, structure, online, interaction, solution, browsing, personal, integration, idea, paper extract, datum source, author, class, agreement, format, world view, fusion process, creator, diary entry, access structure, categorization, categorization scheme, mail, designer], class org.ontoware.text2onto.pom.POMRelation =[play(knowledge, role), serve(management, purpose), take_into(portal, account), allow_for(semantics, knowledge exchange), consist_of(portal architecture, knowledge base), become(document, process), take_in(information, environment), take(information, place)]]


Text2Onto
File Help
Algorithms
ConceptExtraction
TFIDFConceptExtr
InstanceExtraction
ExampleInstanceE
SimilarityExtraction
ContextSimilarityE
ContextExtrac
ConceptClassification
PatternConceptCla
VerticalRelationsC
WordNetConceptC
InstanceClassification
ContextInstanceC
PatternInstanceCla
RelationExtraction
SubcatRelationExt
Corpus
H:\Corpus\corpus_sw\
Concepts
Subclass-of
Instances
Instance-of
Relations
Similiarity
Domain
Range
Probability
datum information 1.0
contents information 1.0
internet network 1.0
template model 1.0
interoperability quality 1.0
contents content 1.0
template content 1.0
semantics content 1.0
meaning content 1.0
knowledge base content 1.0
[subclass-of(discussion, communication), 1.0]
document communication 0.75
documentation communication 0.6666666666666666
network system 0.6
member part 0.6
report communication 0.5714285714285714
technique knowledge 0.5
technology knowledge 0.5
meaning knowledge 0.5
computing knowledge 0.5
technique method 0.5
technology application 0.5
discussion language 0.5
language communication 0.5
hierarchy organization 0.5
management organization 0.5
technology use 0.5
usage use 0.5
usage practice 0.5
technology content 0.5
annotation content 0.5
gy, process management project, layer, warehouse modeling, representation, meta model, fact, process expert, gloss
ary, factor, experiment, device, modeling world, knowledge management process, interface engine, modeling approach
, student, staff, health insurance company, process modeling, configure, category, uniform, process, iphus, suit,
note, group filespace, label, structure, online, interaction, solution, browsing, personal, integration, idea, pap
er extract, datum source, author, class, agreement, format, world view, fusion process, creator, diary entry, acce
ss structure, categorization, categorization scheme, mail, designer], class org.ontoware.text2onto.pom.POMRelation
=[play(knowledge, role), serve(management, purpose), take_into(portal, account), allow_for(semantics, knowl
edge exchange), consist_of(portal architecture, knowledge base), become(document, process), take_in(informat
ion, environment), take(information, place)]]

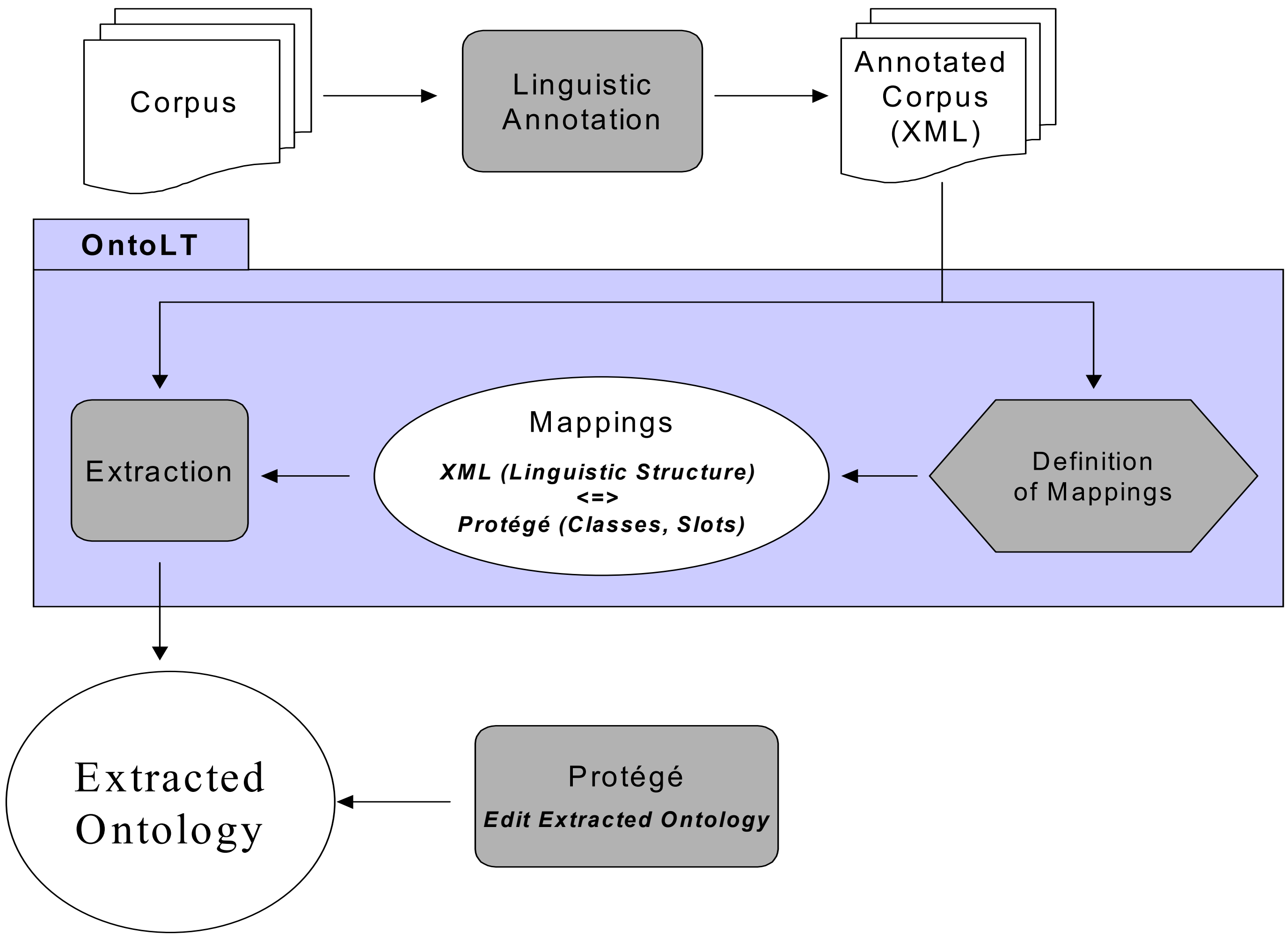
Text2Onto: Data-driven Change Discovery

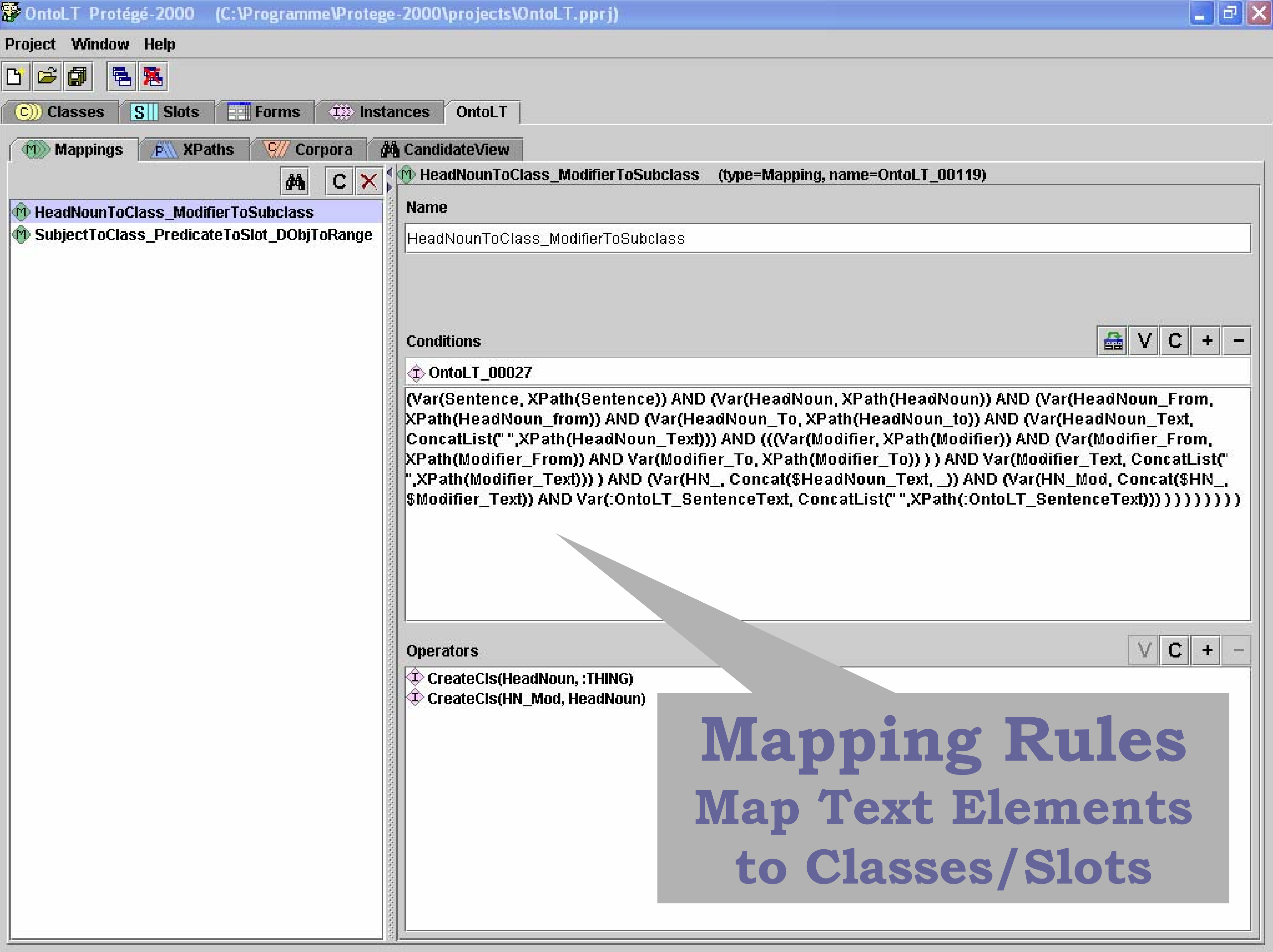


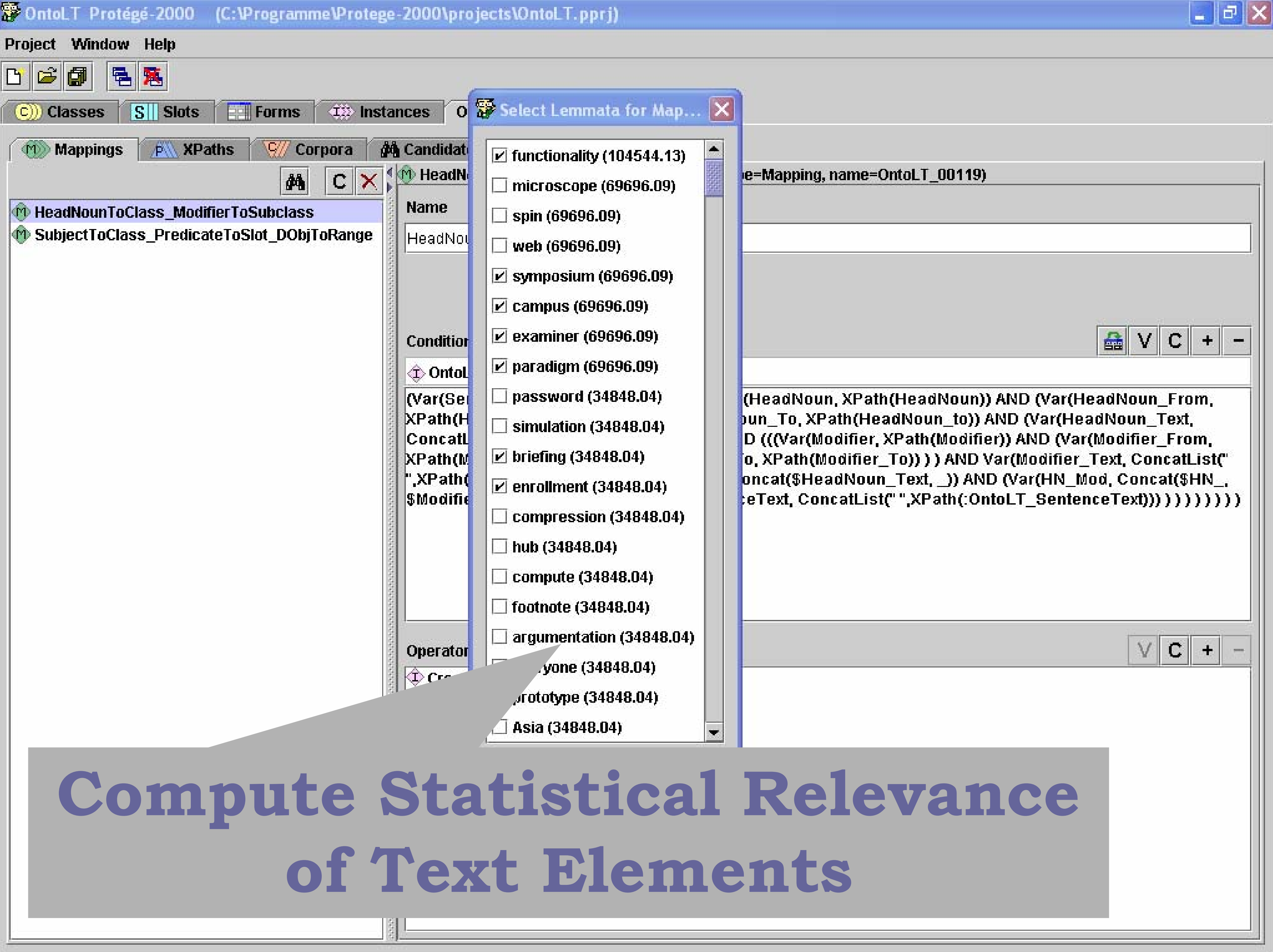
OntoLT (DFKI LT, Saarbrücken)

- Methods:
  - Term extraction by statistical methods (X^2)
  - Definition of linguistic patterns as well as mapping to ontological structures
- Availability: open source (Java, Protégé plugin)
- Link: *http://olp.dfki.de/OntoLT/OntoLT.htm*

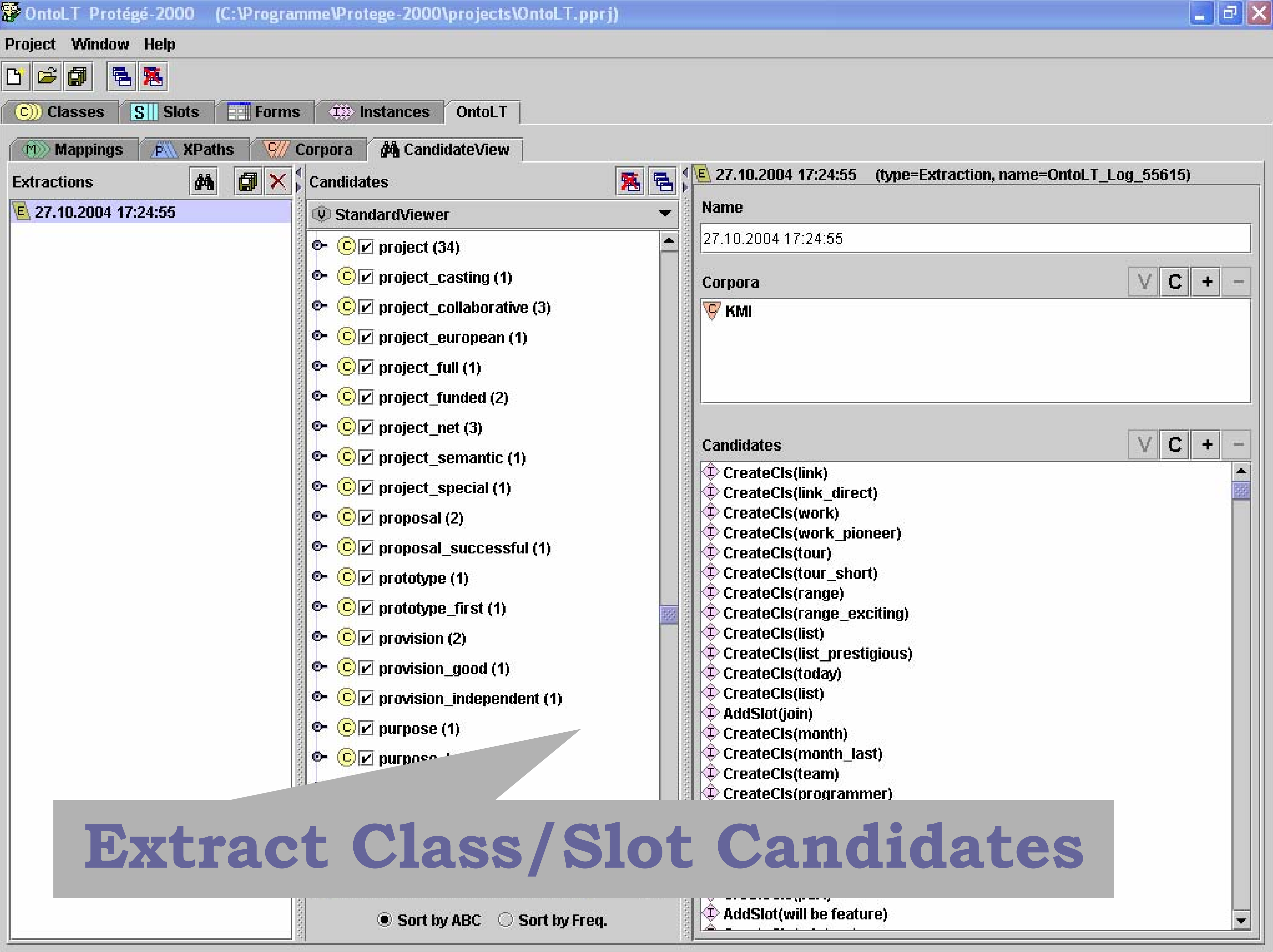
OntoLT: Architecture

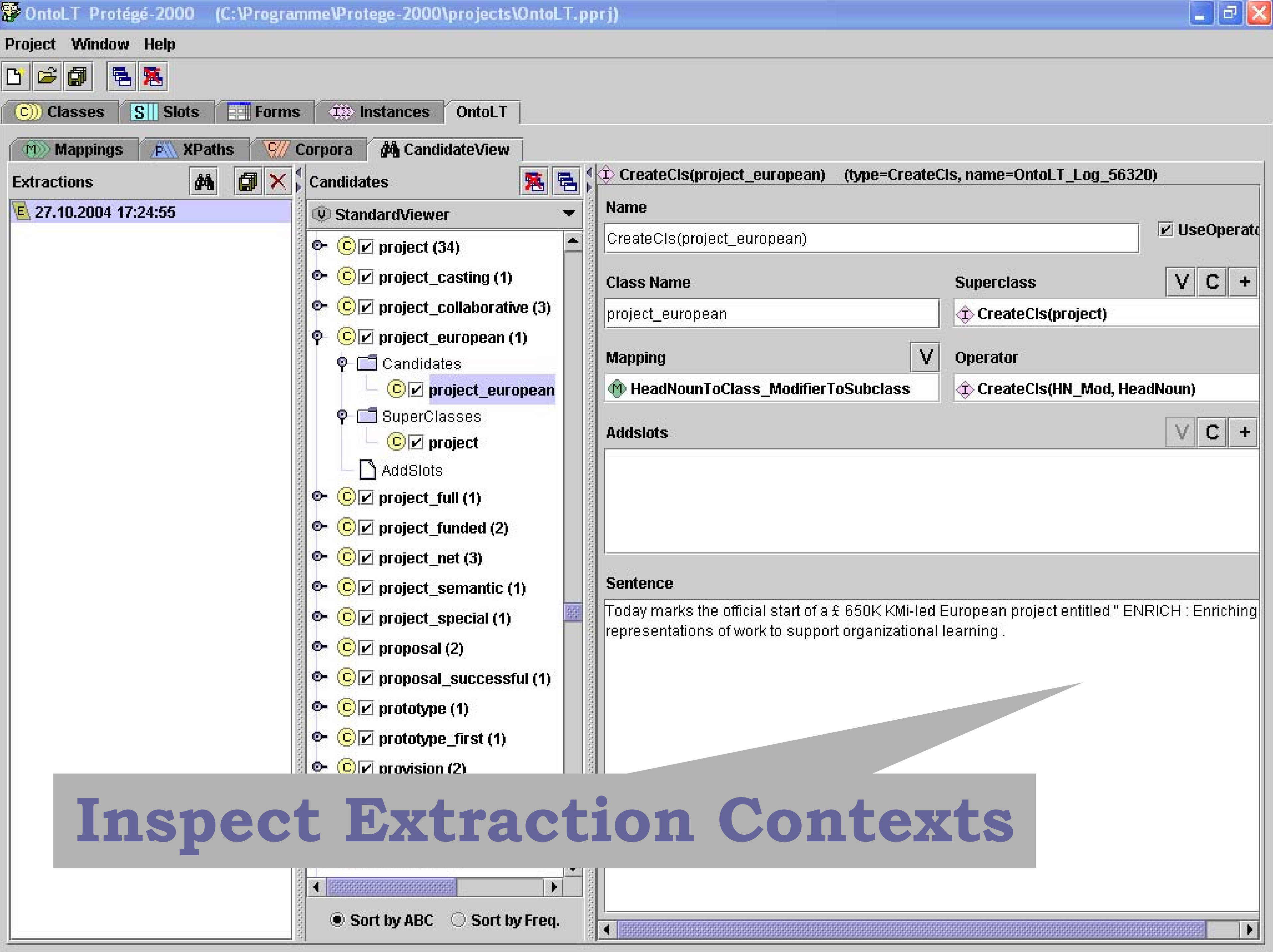


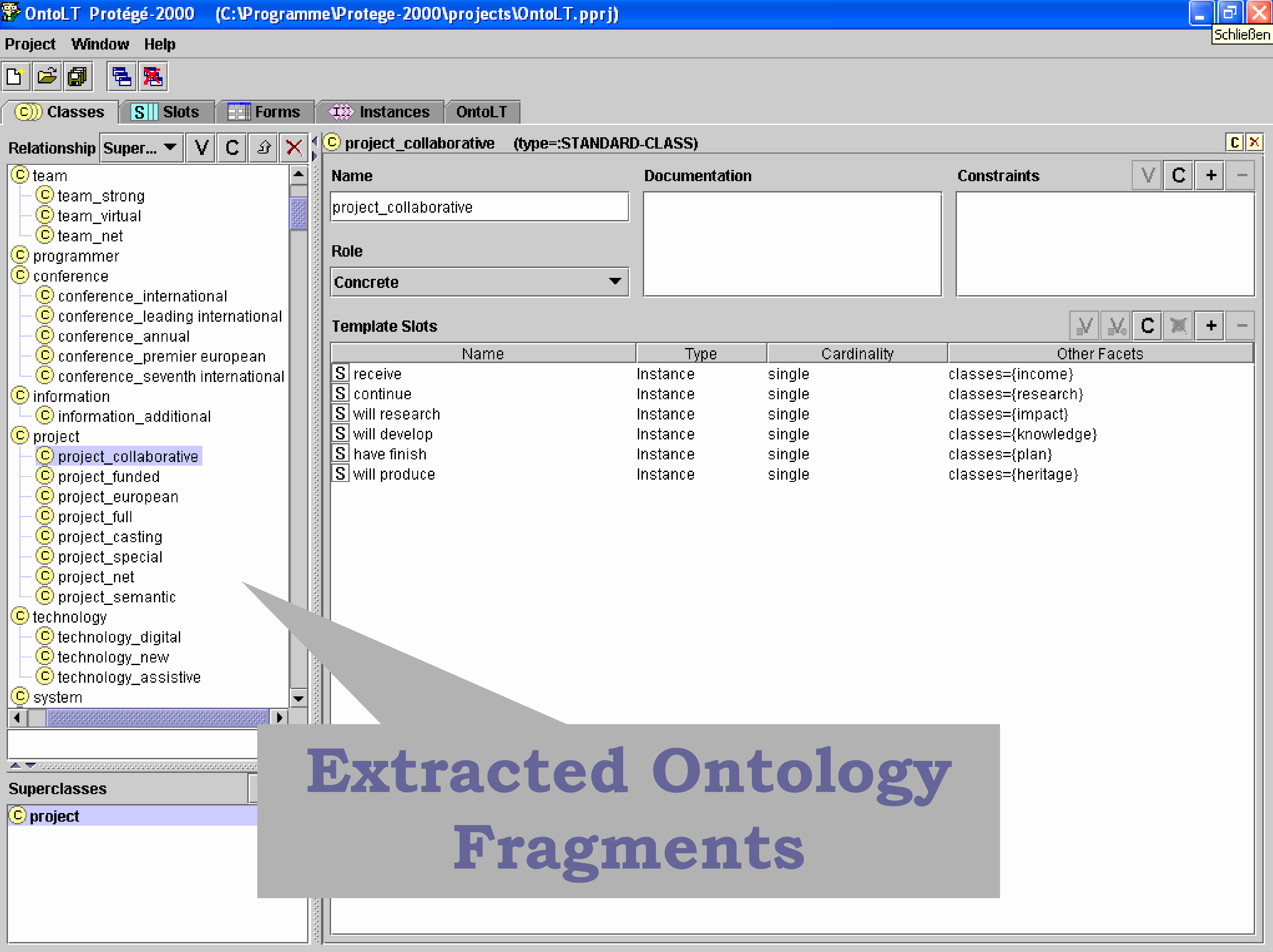

OntoLT Protégé-2000 (C:\Programme\Protege-2000\projects\OntoLT.pprj)
Project Window Help
Classes
Slots
Forms
Instances
OntoLT
Mappings
XPaths
Corpora
CandidateView
HeadNounToClass_ModifierToSubclass
SubjectToClass_PredicateToSlot_DObjToRange
HeadNounToClass_ModifierToSubclass (type=Mapping, name=OntoLT_00119)
Name
HeadNounToClass_ModifierToSubclass
Conditions
OntoLT_00027
(Var(Sentence, XPath(Sentence)) AND (Var(HeadNoun, XPath(HeadNoun)) AND (Var(HeadNoun_From, XPath(HeadNoun_from)) AND (Var(HeadNoun_To, XPath(HeadNoun_to)) AND (Var(HeadNoun_Text, ConcatList(" ",XPath(HeadNoun_Text)) AND (((Var(Modifier, XPath(Modifier)) AND (Var(Modifier_From, XPath(Modifier_From)) AND Var(Modifier_To, XPath(Modifier_To)))) AND Var(Modifier_Text, ConcatList(" ",XPath(Modifier_Text)))) AND (Var(HN_, Concat($HeadNoun_Text, _)) AND (Var(HN_Mod, Concat($HN_, $Modifier_Text)) AND Var(:OntoLT_SentenceText, ConcatList(" ",XPath(:OntoLT_SentenceText))))))))))
Operators
CreateCls(HeadNoun, :THING)
CreateCls(HN_Mod, HeadNoun)
Mapping Rules
Map Text Elements
to Classes/Slots


OntoLT Protégé-2000 (C:\Programme\Protege-2000\projects\OntoLT.pprj)
Project Window Help
Classes
Slots
Forms
Instances
Mappings
XPaths
Corpora
HeadNounToClass_ModifierToSubclass
SubjectToClass_PredicateToSlot_DObjToRange
Name
Condition
Operator
Select Lemmata for Map...
functionality (104544.13)
microscope (69696.09)
spin (69696.09)
web (69696.09)
symposium (69696.09)
campus (69696.09)
examiner (69696.09)
paradigm (69696.09)
password (34848.04)
simulation (34848.04)
briefing (34848.04)
enrollment (34848.04)
compression (34848.04)
hub (34848.04)
compute (34848.04)
footnote (34848.04)
argumentation (34848.04)
Asia (34848.04)
Compute Statistical Relevance
of Text Elements

Extract Class/Slot Candidates




OntoLT Protégé-2000 (C:\Programme\Protege-2000\projects\OntoLT.pprj)
Project Window Help
Classes Slots Forms Instances OntoLT
Mappings XPaths Corpora CandidateView
Extractions
27.10.2004 17:24:55
Candidates
StandardViewer
project (34)
project_casting (1)
project_collaborative (3)
project_european (1)
Candidates
project_european
SuperClasses
project
AddSlots
project_full (1)
project_funded (2)
project_net (3)
project_semantic (1)
project_special (1)
proposal (2)
proposal_successful (1)
prototype (1)
prototype_first (1)
provision (2)
Sort by ABC
Sort by Freq.
CreateCls(project_european) (type=CreateCls, name=OntoLT_Log_56320)
Name
CreateCls(project_european)
UseOperato
Class Name
project_european
Superclass
CreateCls(project)
Mapping
HeadNounToClass_ModifierToSubclass
Operator
CreateCls(HN_Mod, HeadNoun)
Addslots
Sentence
Today marks the official start of a £ 650K KMi-led European project entitled " ENRICH : Enriching representations of work to support organizational learning .
Inspect Extraction Contexts


OntoLT Protégé-2000 (C:\Programme\Protege-2000\projects\OntoLT.pprj)
Schließen
Project Window Help
Classes
Slots
Forms
Instances
OntoLT
Relationship Super...
team
team_strong
team_virtual
team_net
programmer
conference
conference_international
conference_leading international
conference_annual
conference_premier european
conference_seventh international
information
information_additional
project
project_collaborative
project_funded
project_european
project_full
project_casting
project_special
project_net
project_semantic
technology
technology_digital
technology_new
technology_assistive
system
Superclasses
project
project_collaborative (type=:STANDARD-CLASS)
Name
project_collaborative
Documentation
Constraints
Role
Concrete
Template Slots
Name
Type
Cardinality
Other Facets
receive Instance single classes={income}
continue Instance single classes={research}
will research Instance single classes={impact}
will develop Instance single classes={knowledge}
have finish Instance single classes={plan}
will produce Instance single classes={heritage}
Extracted Ontology Fragments

OntoLearn (Department of Computer Science, University „La Sapienza“, Rome)

- Methods
  - Interpretation of compounds by compositional interpretation
    - Disambiguation of terms with respect to WordNet
    - Identify relation between terms in a compound
  - Gloss generation

- Availability: soon online version
- Link: *http://www.dsi.uniroma1.it/~navigli/*

ASIUM (Faure and Nedellec)

- Methods
  - Taxonomy induction by bottom-up clustering of words on the basis of syntactic dependencies
  - Learning of subcategorization frames with respect to the induced taxonomy
- Other features.
  - Cooperative validation of the clusters by the user
- Availability: Unix
- sent on request (contact *claire.nedellec@jouy.inra.fr*)

Mo'K Workbench (Bison et al.)

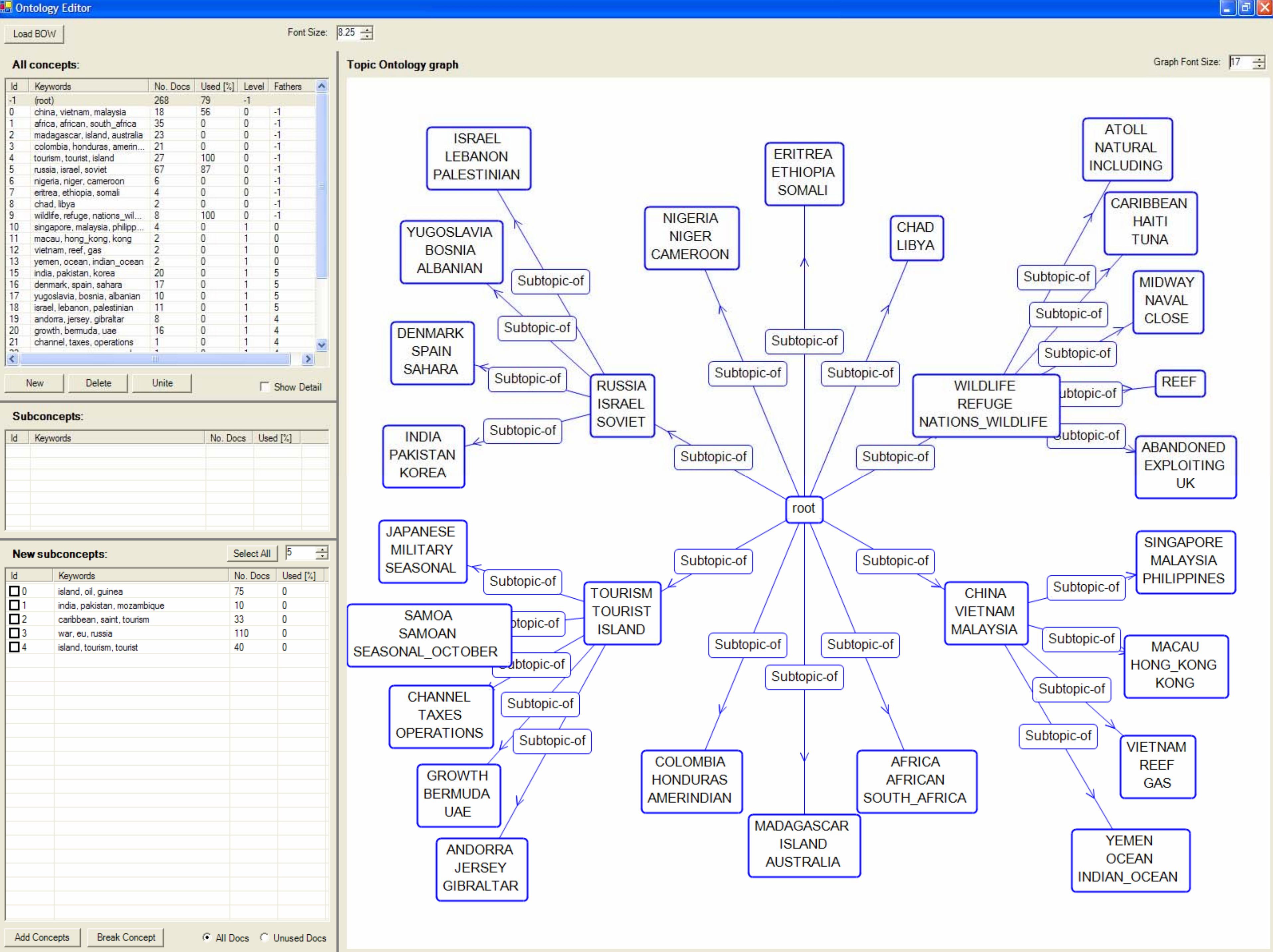
- Methods
  - Workbench allowing to vary:
    - Features describing a word
    - Thresholds
    - similarity/distance measure

- Availability: Mac OS with Mac Common Lisp
- sent on request (contact *gilles.bisson@imag.fr*)

OntoGen (Jožef Stefan Institute)

- Software for semi-automatic generation of ontologies from documents
  - …concepts are proposed by system using LSI/SVD and/or Clustering
  - …concepts are described by terms which best separate concept documents from the rest using Linear Support Vector Machine (SVM)

- Availability: open source (C++, .NET)
- Link: *http://www.textmining.net*
  http://www.sekt-project.com



Ontology Editor
Load BOW
Font Size: 8.25
All concepts:
Id | Keywords | No. Docs | Used [%] | Level | Fathers
-1 (root) 268 79 -1
0 china, vietnam, malaysia 18 56 0 -1
1 africa, african, south_africa 35 0 0 -1
2 madagascar, island, australia 23 0 0 -1
3 colombia, honduras, amerin... 21 0 0 -1
4 tourism, tourist, island 27 100 0 -1
5 russia, israel, soviet 67 87 0 -1
6 nigeria, niger, cameroon 6 0 0 -1
7 eritrea, ethiopia, somali 4 0 0 -1
8 chad, libya 2 0 0 -1
9 wildlife, refuge, nations_wil... 8 100 0 -1
10 singapore, malaysia, philipp... 4 0 1 0
11 macau, hong_kong, kong 2 0 1 0
12 vietnam, reef, gas 2 0 1 0
13 yemen, ocean, indian_ocean 2 0 1 0
15 india, pakistan, korea 20 0 1 5
16 denmark, spain, sahara 17 0 1 5
17 yugoslavia, bosnia, albanian 10 0 1 5
18 israel, lebanon, palestinian 11 0 1 5
19 andorra, jersey, gibraltar 8 0 1 4
20 growth, bermuda, uae 16 0 1 4
21 channel, taxes, operations 1 0 1 4
New
Delete
Unite
Show Detail
Subconcepts:
Id | Keywords | No. Docs | Used [%]
New subconcepts:
Select All
5
Id | Keywords | No. Docs | Used [%]
0 island, oil, guinea 75 0
1 india, pakistan, mozambique 10 0
2 caribbean, saint, tourism 33 0
3 war, eu, russia 110 0
4 island, tourism, tourist 40 0
Add Concepts
Break Concept
All Docs
Unused Docs
Topic Ontology graph
Graph Font Size: 17
root
RUSSIA ISRAEL SOVIET
ISRAEL LEBANON PALESTINIAN
YUGOSLAVIA BOSNIA ALBANIAN
DENMARK SPAIN SAHARA
INDIA PAKISTAN KOREA
NIGERIA NIGER CAMEROON
ERITREA ETHIOPIA SOMALI
CHAD LIBYA
WILDLIFE REFUGE NATIONS_WILDLIFE
ATOLL NATURAL INCLUDING
CARIBBEAN HAITI TUNA
MIDWAY NAVAL CLOSE
REEF
ABANDONED EXPLOITING UK
TOURISM TOURIST ISLAND
JAPANESE MILITARY SEASONAL
SAMOA SAMOAN SEASONAL_OCTOBER
CHANNEL TAXES OPERATIONS
GROWTH BERMUDA UAE
ANDORRA JERSEY GIBRALTAR
COLOMBIA HONDURAS AMERINDIAN
MADAGASCAR ISLAND AUSTRALIA
AFRICA AFRICAN SOUTH_AFRICA
CHINA VIETNAM MALAYSIA
SINGAPORE MALAYSIA PHILIPPINES
MACAU HONG_KONG KONG
VIETNAM REEF GAS
YEMEN OCEAN INDIAN_OCEAN
Subtopic-of

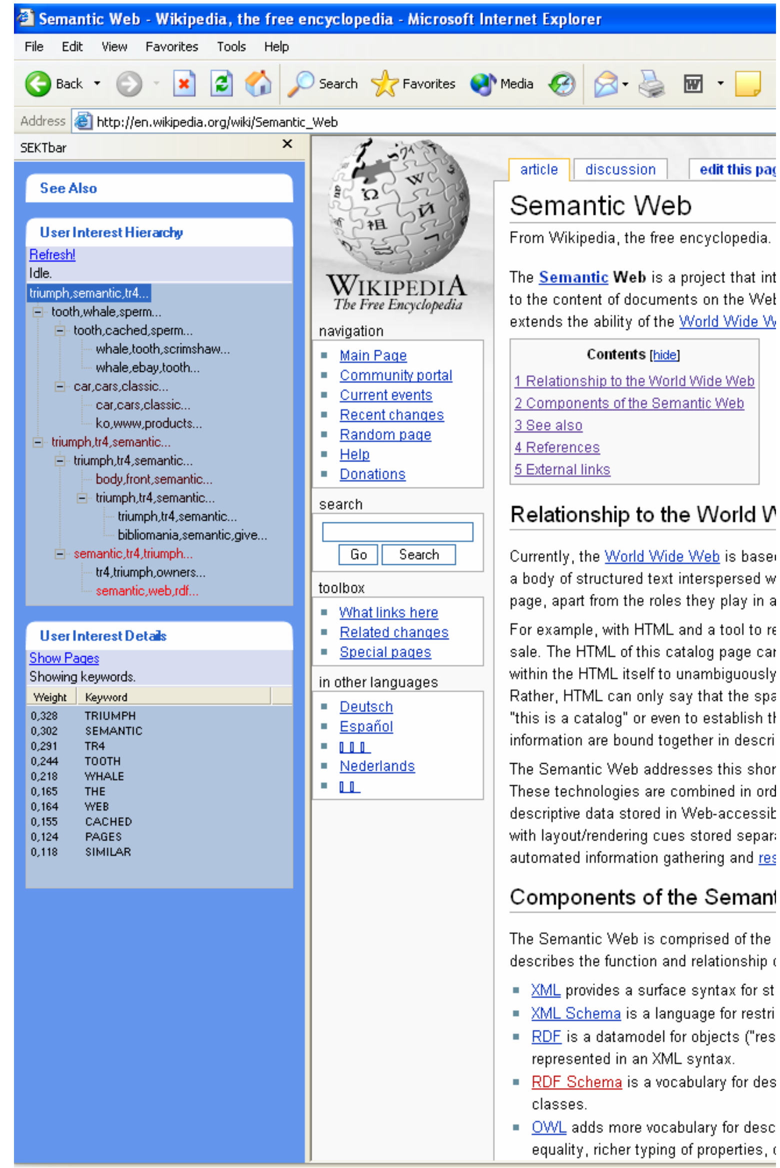
SEKTbar: User profiling

Jožef Stefan Institute

- A Web-based user profile is automatically generated while the user is browsing the Web.
  - It is represented in the form of a user-interest-hierarchy (UIH).
  - The root node holds the user's general interest, while leaves hold more specific interests
  - UIH is generated by using hierarchical k-means clustering algorithm
  - Nodes of current interest are determined by comparing UIH node centroids to the centroid computed out of the *m* most recently visited pages.
- The user profile is visualized on the SEKTbar (Internet Explorer Toolbar)
  - The user can select a node in the hierarchy to see its specific keywords and associated pages (documents)
- Availability: open source (C++, .NET)
- Link: *http://www.textmining.net* *http://www.sekt-project.com*

SEKTbar Example

- The screenshot shows the profile visualization after looking at three distinct topics:
  - "whale tooth"
  - "Triumph TR4"
  - "semantic web"



References

[Abecker and van Elst, 2004] - A. Abecker, L. van Elst. Ontologies for Knowledge Management. In: S. Staab and R. Studer (Eds.), Handbook on Ontologies, pp. 435-454, Springer, 2004.

[Abecker et al. 1997] - A. Abecker, S. Decker, K. Hinkelmann, U. Reimer. In: Proceedings of the International Workshop on Knowledge-Based Systems for Knowledge Management in Enterprises at the German AI Conference (KI-97), 1997.

[Agichtein and Gravano, 2000] - E. Agichtein, L. Gravano, Snowball: Extracting Relations from Large Plain-Text Collections. In: Proceedings of the 5th ACM International Conference on Digital Libraries (ACM DL), pp. 85-94, 2000.

[Agirre and Rigau 1996] - E. Agirre, G. Rigau. Word sense disambiguation using conceptual density. In: Proceedings of the International Conference on Computational Linguistics (COLING'96), pp. 16-22, 1996.

[Ahmad et al. 2003] - K. Ahmad, M. Tariq, B. Vrusias, C. Handy. Corpus-Based Thesaurus Construction for Image Retrieval in Specialist Domains. In: Proceedings of the 25th European Conference on Advances in Information Retrieval (ECIR), pp. 502-510, 2003.

[Alfonseca and Manandhar, 2002] - E. Alfonseca, S. Manandhar. Extending a Lexical Ontology by a Combination of Distributional Semantics Signatures. In: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2002), pp. 1-7, 2002.

References

[Amann and Fundulaki 1999] - B. Amann, I. Fundulaki. Integrating Ontologies and Thesauri to build RDF Schemas. In: Proceedings of ECDL, 1999.

[Aschoff et al. 2004] - F.-R. Aschoff, F. Schmalhofer, L. van Elst. Knowledge Mediation: A Procedure for the Cooperative Construction of Domain Ontologies. In: Proceedings of the ECAI Workshop on Agent-mediated Knowledge Management (AMKM-2004), pp. 29-38, 2004.

[Beale et al.1995] - S. Beale, S. Nirenburg, K. Mahesh. Semantic Analysis in the Mikrokosmos Machine Translation Project. In: Proceedings of the 2nd Symposium on Natural Language Processing, pp. 297-307, 1995.

[Bisson et al. 2000] - G. Bisson, C. Nedellec, L. Canamero. Designing clustering methods for ontology building - The Mo'K workbench. In: Proceedings of the ECAI Ontology Learning Workshop, pp. 13-19, 2000.

[Brewster et al. 2004] - C. Brewster, H. Alani, D. Dasmahapatra, Y. Wilks, Data driven ontology evaluation. In: Proceedings of International Conference on Language Resources and Evaluation (LREC), pp. 26–28, 2004.

[Burton-Jones et al. 2004] – A. Burton-Jones, V.C. Storey, V. Sugumaran, P. Ahluwalia, A semiotic metrics suite for assessing the quality of ontologies. Data and Knowledge Engineering, 2004.

References

[Buitelaar, Sintek 2004] – P. Buitelaar, M. Sintek. OntoLT Version 1.0: Middleware for Ontology Extraction from Text. In: Proceedings. of the Demo Session at the International Semantic Web Conference (ISWC), 2004.

[Buitelaar et al. 2004b] – P. Buitelaar, D. Olejnik, M. Hutanu, A. Schutz, T. Declerck, M. Sintek Towards Ontology Engineering Based on Linguistic Analysis. In: Proceedings of LREC, 2004.

[Buitelaar et al . 2004c] - P. Buitelaar, D. Olejnik, M. Sintek. A Protégé Plug-In for Ontology Extraction from Text Based on Linguistic Analysis. In: Proceedings of the 1st European Semantic Web Symposium (ESWS), 2004.

Buitelaar et al., 2005] – P. Buitelaar, M. Sintek, M. Kiesel. Integrated Representation of Domain Knowledge and Multilingual, Multimedia Content Features for Cross-Lingual, Cross-Media Semantic Web Applications, In Proceedings of ISWC, 2005.

[Caraballo 1999] – S.A. Caraballo. Automatic construction of a hypernym-labeled noun hierarchy from text. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp. 120-126, 1999.

[Cederberg and Widdows 2003] – S. Cederberg, D. Widdows. Using LSA and Noun Coordination Information to Improve the Precision and Recall of Automatic Hyponymy Extraction. In: Proceedings of the Conference on Natural Language Learning (CoNNL), 2003.

References

[Charniak, Berland 1999] - E. Charniak, M. Berland. Finding parts in very large corpora. In: Proceedings of the 37th Annual Meeting of the ACL, pp. 57-64, 1999.

[Chawathe et al. 1996] – S.S. Chawathe, A. Rajaraman, H. Garcia-Molina, J. Widom. Change Detection in Hierarchically Structured Information. In *Proceedings of the ACM SIGMOD Conference*, pp. 493–504, 1996.

[Cimiano et al. 2004] - P. Cimiano, S. Handschuh, S. Staab. Towards the Self-Annotating Web. IN: Proceedings of the 13th World Wide Web Conference, pp. 462-471, 2004.

[Cimiano et al. 2004b] – P. Cimiano, A. Hotho, S. Staab. Comparing Conceptual, Partitional and Agglomerative Clustering for Learning Taxonomies from Text In: Proceedings of the European Conference on Artificial Intelligence (ECAI'04), pp. 435-439. IOS Press, 2004.

[Cimiano and Staab 2004] - P. Cimiano, S. Staab. Learning by Googling, SIGKDD Explorations, 6(2), 2004.

[Cimiano et al. 2005] - P. Cimiano, G. Ladwig, S. Staab. Gimme, The Context: Context-driven automatic semantic annotation with C-PANKOW, IN: Proceedings of the 14th World Wide Web Conference, 2005.

[Cimiano et al. 2005b] - P. Cimiano, L. Schmidt-Thieme, A. Pivk, S. Staab, Learning Taxonomic Relations from Heterogeneous Evidence, Ontology Learning from Text: Methods, Applications and Evaluation, IOS Press, pp. 59-73, 2005.

References

[Cimiano et al. 2005c] – P. Cimiano and S. Staab, Learning Concept Hierarchies from Text with a Guided Agglomerative Clustering Algorithm. In: Proceedings of the ICML 2005 Workshop on Learning and Extending Lexical Ontologies with Machine Learning Methods. 2005.

[Cimiano and Wenderoth 2005] - P. Cimiano, J. Wenderoth, Automatically Learning Qualia Structures from the Web. In: Proceedings of the ACL Workshop on Deep Lexical Acquisition, pp. 28-37, 2005.

[Ciramita et al. 2005] - M. Ciramita, A. Gangemi, E. Ratsch, J. Saric, I. Rojas. Unsupervised Learning of Semantic Relations between Concepts of a Molecular Biology Ontology. In. Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), 2005.

[Clark and Weir 2002] - S. Clark, D.J. Weir. Class-Based Probability Estimation Using a Semantic Hierarchy. Computational Linguistics, 28(2), pp. 187-206, 2002.

[Cleuziou et al. 2004] - G. Cleuziou, L. Martin, C. Vrain. PoBOC: An Overlapping Clustering Algorithm, Application to Rule-Based Classification and Textual Data. In: Proceedings of the European Conference on Artificial Intelligence (ECAI), pp. 440-444, 2004.

References

[Copestake et al.] - Copestake, A., B. Jones, A. Sanfilippo, H. Rodriguez, P. Vossen, S. Montemagni, E. Marinai. Multilingual Lexical Representation. ESPRIT BRA-3030 ACQUILEX - WP No. 043.

[Decker et al. 1997] - S. Decker, M. Daniel, M. Erdmann, R. Studer. An Enterprise Reference Scheme for Integrating Model Based Knowledge Engineering and Enterprise Modeling. In Proceedings of EKAW, 1997.

[Decker et al. 1999] - S. Decker, M. Erdmann, D. Fensel, R. Studer}. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information, In. R. Meersman and Z. Tari and S. Stevens (eds.), Database Semantics: Semantic Issues in Multimedia Systems, Kluwer Academic Publishers, 1999.

[Deutscher Wortschatz] - http://wortschatz.uni-leipzig.de/

[Ding et al. 2004] – L. Ding, T. Finin, A. Joshi and R. Pan, R.S. Cost, Y. Peng, P. Reddivari, V. Doshi, J. Sachs. Swoogle: A search and metadata engine for the semantic web. In: *Proceedings 13th ACM Conference on Information and Knowledge Management*, pp. 652–659, 2004.

[Dorow and Widdows 2003] – B. Dorow, D. Widdows. Discovering Corpus-Specific Word Senses. In: Proceedings of EACL, pp. 79-82, 2003.

[Downey et al. 2004] - D. Downey, O. Etzioni, S. Soderland, D. Weld. Learning Text Patterns for Web Information Extraction and Assessment. In: Proceedings of the AAAI Workshop on Adaptive Text Extraction and Mining, 2004.

References

[van Elst et al. 2003]- L. van Elst, V. Dignum, A. Abecker (Eds.): Agent Mediated Knowledge Management, International Symposium AMKM 2003, Stanford, CA, USA, 2003.

[van Elst and Abecker 2002] - L. van Elst, A. Abecker. Ontologies for Information Management: Balancing Formality, Stability, and Sharing Scope. Expert Systems with Applications, 23(4):357-366, 2002.

[van Elst and Abecker 2002b] Ludger van Elst, Andreas Abecker. Domain Ontology Agents for Distributed Organizational Memories. In: Rose Dieng-Kuntz and Nada Matta (eds.): Knowledge Management and Organizational Memories. Kluwer, 2002.

[Etzioni et al. 2004] - O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D.S. Weld, A. Yates, Web-Scale Information Extraction in KnowItAll (Preliminary Results), In: Proceedings of the 13th World Wide Web Conference, pp. 100-109, 2004.

[Etzioni et al. 2005] - O. Etzioni, M. Cafarella, D. Downey, A-M. Popescu, T. Shaked, S. Soderland, D.S. Weld, A. Yates, Unsupervised Named-Entity Extraction from the Web: An Experimental Study. Artificial Intelligence, 165(1), pp. 91-134, 2005.

[Faure and Nedellec, 1998] – D. Faure, C. Nedellec. A corpus-based conceptual clustering method for verb frames and ontology acquisition. In: Proceedings of LREC Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Applications, 1998.

References

[Fensel 2001] - D. Fensel, Ontologies: Silver bullet for knowledge management and electronic commerce, Springer, 2001.

[FIPA] Foundations for Intelligence Physical agents (http://www.fipa.org/)

[Fisher 1987] - D. Fisher, Knowledge acquisition via incremental conceptual clustering. Machine Learning 2, pp. 139-172, 1987.

[Firth 1957] - J. Firth, A synopsis of linguistic theory 1930-1955, Longman, Studies in Linguistic Analysis, Philological Society, 1957.

[Fortuna et al., 2005] - B. Fortuna, D. Mladenic, M. Grobelnik Visualization of text document corpus. ACAI 2005 Summer School.

[Fotzo, Gallinari 2004] - H.N. Fotzo, P. Gallinari, Learning Generalization/Specialization Relations between Concepts - Application for Automatically Building Thematic Document Hierarchies, In: Proceedings of RIAO, 2004.

[Fox et al., 1998] - Fox, M. S., Barbuceanu, M., Gruninger, M., Lin, J., An organization ontology for enterprise modeling. In: M. Prietula et al. (eds.), Simulating organizations: Computational models of institutions and groups, AAAI/MIT Press, 1998, pp. 131-152.

[Frantzi and Ananiadou, 1999] – K.T. Frantzi, S. Ananiadou.The C-Value/NC-Value domain independent method for multi-word term extraction. Journal of Natural Language Processing, 6(3):145-179,1999.

[Ganter and Wille 1999] – B. Ganter, R. Wille. Formal Concept Analysis – Mathematical Foundations, Springer Verlag, 1999.

References

[Gasperin et al. 2001] - C. Gasperin, P. Gamallo, A. Agustini, G. Lopes and V. de Lima, Using Syntactic Contexts for Measuring Word Similarity. In: Proceedings of the ESSLLI Workshop on Semantic Knowledge Acquisition and Categorization, 2001.

[Girju et al. 2002] - R. Girju, D. Moldovan, Text Mining for Causal Relations, In: Proceedings of the FLAIRS Conference, pp. 360-364, 2002.

[Gluschko et al. 1999] - R. J. Gluschko and J. M. Tenenebaum and B. Meltzer. An XML Framework for Agent-based E-Commerce. In: Communications of the ACM 42(3):106-114, 1999.

[Gomez-Perez 1994] - Gómez-Pérez. A. Some ideas and examples to evaluate ontologies. Knowledge Systems Laboratory, Stanford University, 1994.

[Gomez-Perez 1996] - Gómez-Pérez. A. Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications*, 11(4):519–529, 1996.

[Grefenstette, 1992] - Grefenstette. Sextant: Exploring unexplored contexts for semantic extraction from syntactic analysis. In: Proceedings of the *30th Annual Meeting of the Association for Computational Linguistics,* Newark, Delaware, 28 June - 2 July 1992.

[Grefenstette 1992] – G. Grefenstette. Evaluation techniques for automatic semantic extraction: Comparing syntactic and window-based approaches. In: Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text, 1992.

[Grefenstette 1998] – G. Grefenstette. Cross-Language Information Retrieval, Kluwer Academic Publishing, 1998.

References

[Gruber 1993] - T.R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, Formal Analysis in Conceptual Analysis and Knowledge Representation, Kluwer, 1993.

[Guarino and Welty 2002] - Guarino, N., Welty, C., Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45(2):61–65, 2002.

[Guarino et al. 1999] - N. Guarino, C. Masolo, G. Vetere. OntoSeek: Content-Based Access to the Web. In: IEEE Intelligent Systems, 14(3), 70--80, 1999.

[Haase and Völker, 2005] - P. Haase, J. Völker, Ontology Learning and Reasoning -- Dealing with Uncertainty and Inconsistency. In: Proceedings of the Workshop on Uncertainty Reasoning for the Semantic Web (URSW), 2005.

[Hartmann et al. 2005] – J. Hartmann, P. Spyns, A. Giboin, D. Maynard, R. Cuel, M.C. Suárez-Figueroa, Y. Sure. Methods for ontology evaluation. KnowledgeWeb (EU-IST Network of Excellence IST-2004-507482 KWEB), Deliverable D1.2.3, January 2005.

[Harris 1968] - Z.S. Harris. Mathematical Structures of Language. Wiley, 1968.

[Hearst 1992] - M.A. Hearst, Automatic Acquisition of Hyponyms from Large Text Corpora. In: Proceedings of the 14th International Conference on Computational Linguistics, pp. 539-545, 1992.

[Hendler 2000] - J. Heflin, J. Hendler. Searching the Web with SHOE, In: Papers from the AAAI Workshop on Artificial Intelligence for Web Search, pp. 35-40, 2000.

References

[Iwanska et al., 2000] - L.M. Iwanska, N. Mata, K. Kruger. Fully Automatic Acquisition of Taxonomic Knowledge from Large Corpora of Texts. Natural Language Processing and Knowledge Processing, 335--345, MIT/AAAI Press, 2000.

[Kashyap 1999] - V. Kashyap. Design and Creation of Ontologies for Environmental Information Retrieval. Proceedings of the 11th European Workshop on Knowledge Acquisistion, Modeling,and Management (EKAW), 1999.

[Kavalec and Svatek, 2005] – M. Kavalec, V. Svatek. A Study on Automated Relation Labelling. In Ontology Learning. In: P.Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning and Population from Text: Methods, Evaluation and Applications, IOS Press, 2005.

[Kesseler 1996] - M. Kesseler. A Schema Based Approach to HTML Authoring. In: World Wide Web Journal 96(1), O'Reilly, 1996.

[Lee 1999] – L. Lee. Measures of Distributional Similarity. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp- 25-32, 1999.

[Liddy, 1994] – E.D. Liddy, W. Pail, E.S. Yu, M. McKenna. Document Retrieval Using Linguistic Knowledge. In Proceedings of RIAO 94, pp. 106-114, 1994.

[Lin and Pantel 2001] - D. Lin, P. Pantel, DIRT - Discovery of Inference Rules from Text. In: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 323--328, 2001.

[Lozano-Tello and Gomez-Perez 2004] – A. Lozano-Tello, A. Gómez-Pérez, Ontometric: A method to choose the appropriate ontology. *Journal of Database Management*, 15(2):1–18, 2004.

References

[Maedche 2002] – A. Maedche. Ontology Learning for the Semantic Web. Kluwer Academic Publishers, 2002.

[Maedche and Staab 2002] – A. Maedche, S. Staab, Measuring similarity between ontologies. In: *Proceedings of the 13th Conference on Information and Knowledge Management* (EKAW), 2002.

[Maedche and Staab, 2000] – A. Maedche, S. Staab. Semi-automatic Engineering of Ontologies from Text. In: Proceedings of the 12th International Conference on Software Engineering and Knowledge Engineering, 2000.

[Maedche et al. 2002] - A. Maedche, G. Neumann, S. Staab. Bootstrapping an Ontology-Based Information Extraction System, Studies in Fuzziness and Soft Computing, Intelligent Exploration of the Web, Springer, 2002.

[Maedche et al. 2002] - A. Maedche and V. Pekar and S. Staab. Ontology Learning Part One - On Discovering Taxonomic Relations from the Web. In: Web Intelligence, pp. 301-322, Springer, 2002.

[Markert et al. 2003] - K. Markert and N. Modjeska and M. Nissim, Using the Web for Nominal Anaphora Resolution. In: Proceedings of the EACL Workshop on the Computational Treatment of Anaphora, 2003.

References

[Martin and Eklund 2000] – Ph. Martin and P. Eklund. *Knowledge Indexation and Retrieval and the Word Wide Web.* In: IEEE Intelligent Systems, Special Issue "Knowledge Management and Knowledge Distribution over the Internet", 2000.

[Mena and Kashyap 1998] - E. Mena, V. Kashyap, A Illarramendi, A. Sheth. Domain Specific Ontologies for Semantic Information Brokering on the Global Information Infrastructure. In: Proceedings of FOIS, 1998.

[Mulholland et al. 2001] – P. Mulholland, Z. Zdrahal, J. Domingue, M. Hatala, A. Bernardi. A Methodological Approach to Supporting Organizational Learning. International Journal of Human-Computer Studies, 55 (3), 337-367, 2001.

[Navigli and Velardi, 2004] - R. Navigli, P. Velardi. Learning Domain Ontologies from Document Warehouses and Dedicated Websites, Computational Linguistics (30-2), MIT Press, , 2004.

[Nirenburg and Raskin, 2004] – S. Nirenburg and V. Raskin. Ontological Semantics SERIES: Language, Speech, and Communication, MIT Press, 2004.

[Ogden and Richards, 1923] – C.K. Ogden, I. A. Richards. The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism. 8th ed. 1923. Reprint, New York: Harcourt Brace Jovanovich, 1946.

[OnToAgents] http://www-db.stanford.edu/OntoAgents/

[Patel et al. 2004] - Patel, C., Supekar, K., Lee, Y., Park, E. K., OntoKhoj: a semantic web portal for ontology searching, ranking and classification. In: Proc. of the 5th ACM Intl. Workshop on Web Information and Data Management, pp. 58–61. 2004.

References

[Pantel and Lin 2003] - P. Pantel, D. Lin, Automatically Discovering Word Senses. In: Proceedings of HLT-NAACL, 2003.

[Poesio et al. 2002] - M. Poesio, T. Ishikawa, S. Schulte im Walde, R. Viera. Acquiring Lexical Knowledge for Anaphora Resolution. In: Proceedings of the 3rd Conference on Language Resources and Evaluation (LREC), 2002.

[Poesio and Almuhareb 2005] - M. Poesio, A. Almuhareb. Identifying Concept Attributes Using A Classifier. In: Proceedings of the ACL Workshop on Deep Lexical Acquisition, pp. 18-27, 2005.

[Porzel, Malaka 2004] – R. Porzel, R. Malaka, A task-based approach for ontology evaluation. In: *Proceedings of the ECAI Workshop on Ontology Learning and Population*, pp. 9–16, 2004.

[Priss, in preparation] - U. Priss, Formal Concept Analysis in Information Science, Annual Review of Information Science and Technology, Vol. 40, in preparation.

[Probst et al. 1999] – G. Probst, S. Raub, Steffen, K. Romhardt. Wissen managen. Wie Unternehmen ihre wertvollste Ressource optimal nutzen. Frankfurt am Main, 1999.

[Ravichandran et al. 2005] - D. Ravichandran, P. Pantel, E. Hovy. Randomized Algorithms and NLP: Using Locality Sensitive Hash Functions for High Speed Noun Clustering. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2005.

References

[Reinberger et al., 2004] - M.-L. Reinberger, P. Spyns, A.J. Pretorius, and W. Daelemans, Automatic initiation of an ontology, in R. Meersman, Z. Tari et al. (eds.), On the Move to Meaningful Internet Systems, LNCS 3290 , Springer, 600–617, 2004.

[Riloff, 1993] - E. Riloff, W. Lehnert. Automated Dictionary Construction for Information Extraction from Text. In: *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications*. IEEE Computer Society Press. pp. 93–99, 2003.

[Rinaldi et al., 2005] - Fabio Rinaldi, Elia Yuste, Gerold Schneider, Michael Hess, David Roussel. Exploiting Technical Terminology for Knowledge Management. In: P.Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning and Population, IOS Press, 2005.

[Resnik 1993] P. Resnik. Selection and Information: A Class-Based Approach to Lexical Relationships. PhD Thesis, University of Pennsylvania, 1993.

[Ribas 95] F. Ribas. On learning more appropriate selectional restrictions. In: Proceedings of the 7th Conference of the European chapter of the Association for Computational Linguistics (EACL), pp. 112-118, 1995.

[Sabou, 2005] – Marta Sabou Learning Web Service Ontologies: an Automatic Extraction Method and its EvaluationIn Ontology Learning. In: P.Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning and Population, IOS Press, 2005.

[Sanchez, Moreno, 2005] - D. Sanchez and A. Moreno, Web-scale taxonomy learning, In: Proceedings of the ICML Workshop on Extending and Learning Lexical Ontologies using Machine Learning, 2005.

References

[Saussure 1916] – Ferdinand de Saussure. Cours de linguistique générale. Ed. Charles Bally and Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot, 1916.

[Schank et al. 1973] - R. Schank, N. Goldman, C. Rieger, and C. Riesbeck. MARGIE: Memory Analysis Response Generation and Inference on English. In: Proceedings of IJCAI, 1973.

[Schulte im Walde 2000] - S. Schulte im Walde, Clustering Verbs Semantically According to their Alternation Behaviour. In: Proceedings of the 18th International Conference on Computational Linguistics (COLING), pp. 747-753, 2000.

[Schutz and Buitelaar, 2005] – A. Schutz, P. Buitelaar RelExt: A Tool for Relation Extraction in Ontology Extension. In: Proceedings of the 4th International Semantic Web Conference, 2005.

[Schütze 1993] – H. Schütze. Word space, Advances in Neural Information Processing Systems 5, pp. 895-902, 1993.

[Sintek et al. 2004] – M. Sintek, P. Buitelaar, D. Olejnik. A Formalization of Ontology Learning from Text. In: Proceedings of the ISWC Workshop on Evaluation of Ontology-based Tools (EON2004), 2004.

References

[Smith and Poulter 1999] - H. Smith, K. Poulter. Share the Ontology in XML-based Trading Architectures. In: Communications of the ACM 42(3):110-111, 1999.

[Snow et al. 2004] - R. Snow and D. Jurafsky and A.Y. Ng, Learning syntactic patterns for automatic hypernym discovery, In: Proceedings of Advances in Neural Information Processing Systems 17, 2004.

[Soderland et al., 1995] - Soderland, W., D. Fisher, J. Aseltine, and W. Lehnert, "CRYSTAL: Inducing a Conceptual Dictionary". In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI),* pp. 1314-1319, 1995.

[Sowa, J. F. 1984] – *Conceptual Structures: Information Processing in Mind and Machine*. Reading, Massachusetts, Addison-Wesley, 1985.

[Sparck-Jones, 1966/1986] – K. Sparck Jones. Synonymy and Semantic Classification. Edinburgh University Press, Edinburgh, 1966/1986.

[Sparck Jones 1971] – K. Sparck Jones. Automatic Keyword Classification and Information Retrieval, Butterworths, London, 1971.

[Stumme et al. 2003] G. Stumme, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, Y. Sure, R. Volz, V. Zacharias. The Karlsruhe View on Ontologies. Technical Report University of Karlsruhe, Institute AIFB, 2003.

[Sundblad 2002] – H. Sundblad. Automatic Acquisition of Hyponyms from Question Corpora, Proceedings of the ECAI Workshop on Ontology Learning, 2002.

References

[Superkar 2005] – K. Supekar. A peer-review approach for ontology evaluation. In: *Proceedings of the 8th International Protégé Conference*, 2005.

[Sure 2003] – Y. Sure, Methodology. Tools and Case Studies for Ontology based Knowledge Management. PhD Thesis, University of Karlsruhe, Institute AIFB, 2003.

[Sure et al. 2000] - Y. Sure and A. Maedche and S. Staab. Leveraging Corporate Skill Knowledge -- From ProPer to OntoProPer, In: Proceedings of PAKM, pp. 1-9, 2000.

[TOVE 1995] – TOVE: Manual of the Toronto Virtual Enterprise, Department of Industrial Engineering, University of Toronto, 1995.

[Uschold and Grunninger 1996] - M. Uschold, M. Grunninger. Ontologies. Principles, Methods and Applications. *Knowledge Engineering Review* 11, 1996.

[Staab and Schnurr 2000] - S. Staab and H.-P. Schnurr. Smart Task Support through Proactive Access to Organizational Memory. Journal of Knowledge-based Systems}, Elsevier, 2000.

[Uschold et al. 1998] - M. Uschold and M. King and S. Moralee and Y. Zorgios. The Enterprise Ontology, In: Knowledge Engineering Review, 13(1), pp. 31-89, 1998.

[Velardi et al., 2005] - P. Velardi, R. Navigli, A. Cucchiarelli, F. Neri. Evaluation of OntoLearn, a Methodology for Automatic Learning of Domain Ontologies, In: P.Buitelaar, P. Cimiano, B. Magnini (eds.), Ontology Learning and Population, IOS Press, 2005.

References

[Völker et al. 2005] – J. Völker, D. Vrandecic, Y. Sure, Automatic evaluation of ontologies (AEON). In: Proceedings of the 4th International Semantic Web Conference, 2005.

[Yamada and Bladwin 2004] - I. Yamada, T. Baldwin, Automatic Discovery of Telic and Agentive Roles from Corpus Data. In: Proceedings of the The 18th Pacific Asia Conference on Language, Information and Computation (PACLIC 18), 2004.

[Widdows 2003] - D. Widdows. Unsupervised method for developing taxonomies by combining syntactic and statistical information. In: Proceedings of HLT/NAACL, pp. 276-283, 2003.

[Wiederhold 1992] - G. Wiederhold. Mediators in the architecture of future information systems. In: IEEE Computer 25(3):38-49, 1992.

[Witschel 2005] H.F. Witschel. Using decision trees and text mining techniques for extending taxonomies. In: Proceedings of Learning and Extending Lexical Ontologies by using Machine Learning Methods, Workshop at ICML-05, 2005.

[Woods 1973] - W. A. Woods. Progress in natural language understanding: An application to lunar geology. In: Proc. of the AFIPS Conference, pp. 441–450, 1973.