Integrating Ontologies K-Cap 2005

October 2, 2005

Banff, Canada

Benjamin Ashpole, Marc Ehrig,
Jérôme Euzenat, Heiner Stuckenschmidt

knowledgeweb
realizing the semantic web

- Research Presentations
- Ontology Alignment Evaluation Initiative

Topics

- ontology/schema alignment and matching
- ontology/schema mapping and transformation
- ontology/schema merging and integration
- ontology/schema mediation and reconciliation
- reuse of knowledge from disparate sources (text, user input, etc.) for ontology alignment
- automatic and semi-automatic approaches
- mapping languages
- applications for and tools based on alignment
- integration within larger applications
- evaluation approaches
- translation of information between heterogeneous sources

Agenda I

9:00-10:45	Session 1: Research Presentations Akkermans: Reverse Leibniz, and then Bend It Like Beckham Choi, Hatala: Towards Browsing Distant Metadata Using Semantic Signature Chu, Chow, Chen: Semantic Association of Taxonomy-based Standards Using Ontology Ehrig, Euzenat: Relaxed Precision and Recall
11:10-12:00	Session 2: Research Presentations Gasevic, Hatala: Searching Web Resources Using Ontology Mappings Hu, Jian, Qu, Wang: GMO: A Graph Matching for Ontologies

Agenda II

1:30-2:20	Session 3: Research Presentations Lamsfus: Towards Semantic Based Information Exchange and Integration Standards Silva, Maio, Rocha: An approach to ontology mapping negotiation OAEI Euzenat, Stuckenschmidt: Introduction Ehrig: Results of University of Karlsruhe Valtchev: Results of University of Montreal / INRIA
2:20-3:00	Session 4: OAEI Zanobini: Results of IRST Trento Hu: Results of University of Southhampton Nsjian: Results of Southeast University Nanjin Troncy: Results of CNR / Pisa Wrap-Up

Integrating Ontologies K-Cap 2005

October 2, 2005

Banff, Canada

Benjamin Ashpole, Marc Ehrig,
Jérôme Euzenat, Heiner Stuckenschmidt

knowledgeweb
realizing the semantic web

From Leibniz to Beckham: Time Ontology Mappings as PSM

Hans Akkermans

AKMC
&
Free University Amsterdam VUA
Business Informatics

Positioning: Two Traditions in Temporal Reasoning

- Intervals
- Common in logic and AI temporal reasoning
- Cf. Allen's axiomatization, and associated temporal reasoning methods
- Many different modellings of time possible

- Points
- Standard in science, engineering, and math
- Cf. dynamic systems simulation, and analytical & numerical methods
- Many different modellings of time possible
- This work
  - ➔ How does the (continuous) world look like if you are a fully discrete being?
  - ➔ Two ontological perspectives

Problem: ontology mapping between time (point) ontologies

- Prototypical example 1:
- ODE: Ordinary differential equation: $d/dt\ x_t = f(x_t)$
- Continuous time $t \in \Re$
- D = d/dt is the "generator of the infinitesimal time evolution"

- Prototypical example 2:
- Iterated map: $X_{S+1} = f(X_S)$, or $\Delta X_S \equiv X_{S+1} - X_S = g(X_S)$
- Discrete time $S \in \aleph$
- $N = 1 + \Delta$, the "Next" operator is the generator of the discrete-time evolution

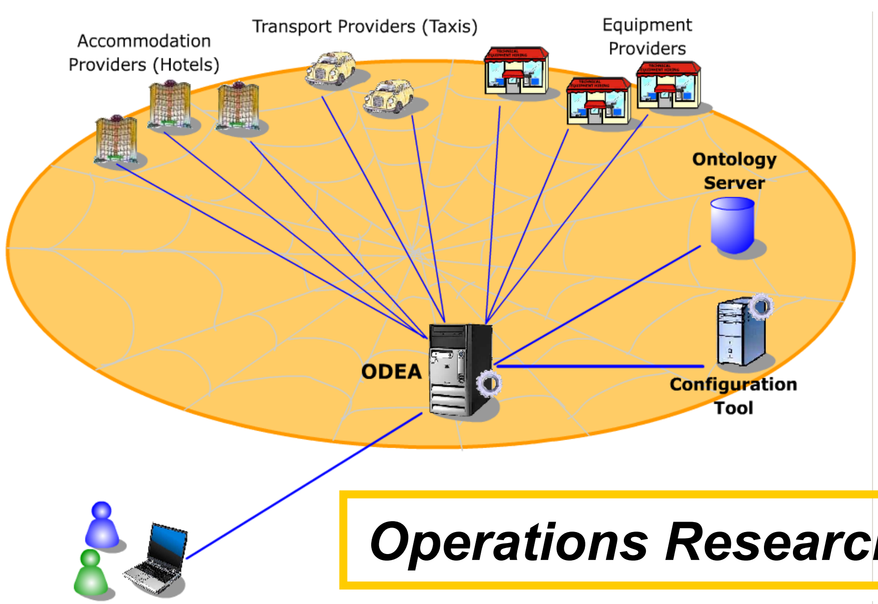
- Key conceptual problem: infinitesimal calculus in discrete space
- In particular: to what concept does the derivative d/dt map?

Motivations (1)

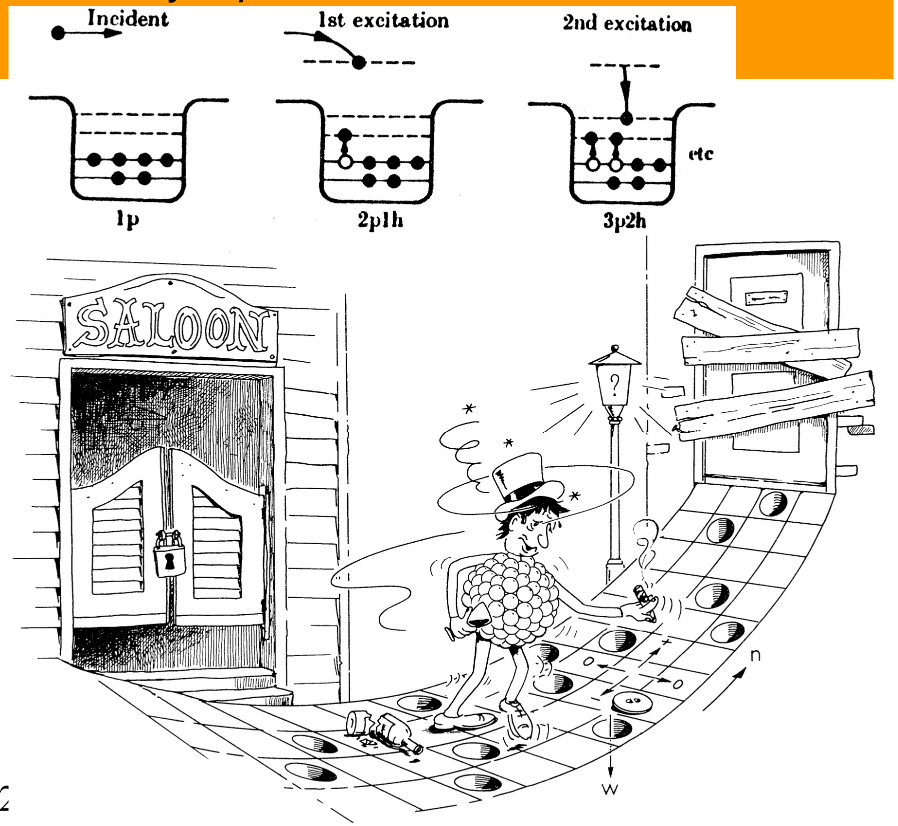
- Empirical
  - ➔ Do alternative but analogous system models actually yield the same empirical predictions?
- Computational
  - ➔ Discrete modellings (if possible) probably have computational advantages since the computer is a discrete machine
- Informational
  - ➔ Alternative modellings may yield different kinds of important information more easily

Motivations (2)

- Conceptual: Many cases exist where dynamic systems can be expressed in alternative models, in both continuous and discrete time
  - ➔ Intuition: underlying ideas are conceptually the same
    - *Are models “really” different or “essentially equivalent”?*
    - *What is “best” ?*



Operations Research

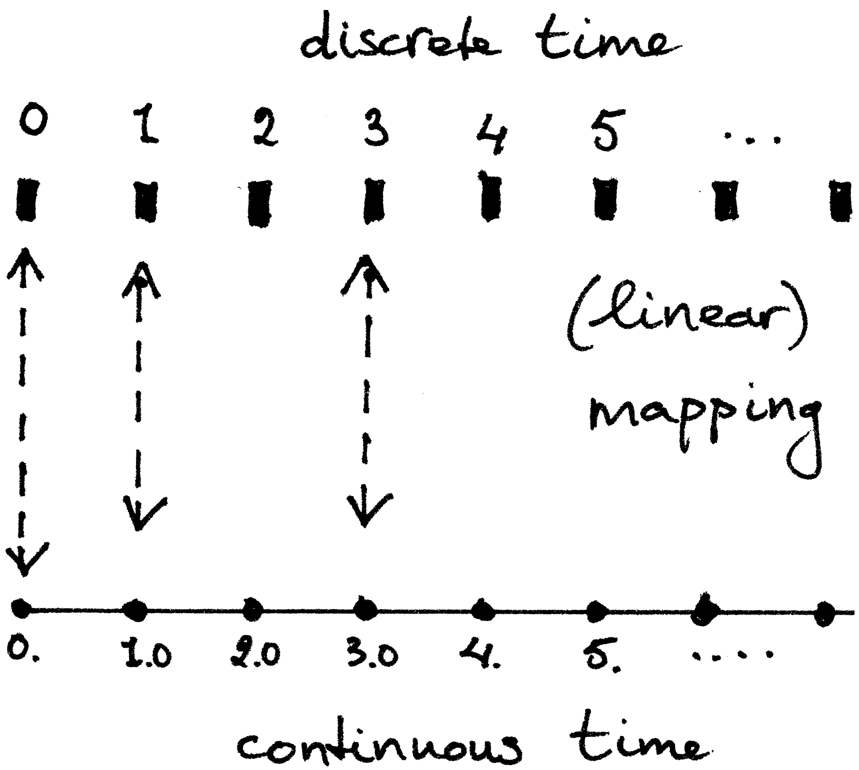


Standard numerical analysis and simulation:

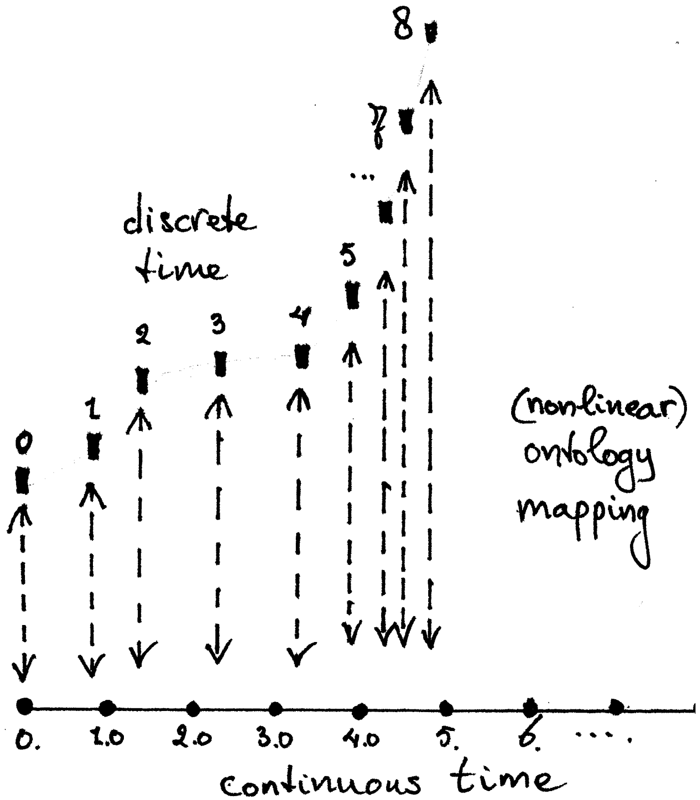
- ODE: $d/dt\ x_t = f(x_t)$
- $d/dt\ x_t \equiv \lim_{\tau\to 0} (x_{t+\tau} - x_t) / \tau$
- Drop $\lim_{\tau\to 0}$ and keep τ finite: $(x_{t+\tau} - x_t) / \tau = f(x_t)$

or

- $\Delta X_S \equiv X_{S+1} - X_S = \tau f(X_S)$
  - *(τ is finite "stepsize")*
- In fact based on linear ontology mapping
- Systematic truncation or discretization error



Nonlinear Ontology Mapping: T Transform



- Discrete events S mapped onto real time t:
  - ➔ Stochastic: discrete points "smeared out" over time
  - ➔ Non-linear, and not one-to-one

$$x_t = \sum_{S=0}^{\infty} P(t,S) \cdot X_S$$

- Specific choice: Poisson probability distribution

$$x_t = \sum_{S=0}^{\infty} \frac{1}{S!} \left(\frac{t}{\tau}\right)^S e^{-t/\tau} X_S$$

= Time spacing of independent events; cf. helpdesk: distribution of calls over time

- Interpreted as **transform** equations

Transform Methods as PSMs

- If a problem is difficult to solve directly, do:
  - ➔ 1. Map problem to new space where it's easier to solve
  - ➔ 2. Solve problem in this new space
  - ➔ 3. Map solution back to original space,
    - *and you are done!*

- Scientific examples: Fourier, Laplace transforms: transform from real time to frequency space
  - ➔ E.g. Differential equations become algebraic equations
- Game examples: mutilated chessboard, Nim
  - ➔ E.g. You win easily!

Key Result

- Derivative D = d/dt maps onto the finite divided difference Δ
- Repeated application of discrete Next operator N = 1 + Δ, gives solution of continuous system
- (This is the *first* computationally practical Taylor-type algorithm)

Some Properties of the T transform

Property No.	Continuous-time function $x_t = T(X_S)$	Discrete-time function $X_S = \mathfrak{T}(x_t)$
I.	1 (constant)	1 (constant)
II.	t	S
III.	t^2	S(S-1)
IV.	t^3	S(S-1)(S-2)
V.	t^n	S! / (S-n+1)!
VI.	e^{At}	$(1 + A)^S$
VII.	$A\, x_t + B\, y_t$	$A\, X_S + B\, Y_S$
VIII.	$d/dt\ x_t$	$\Delta X_S \equiv X_{S+1} - X_S$
IX.	$d^n/dt^n\ x_t$	$\Delta^n X_S$
X.	$f_t \equiv y_t \times x_t$	$F_S = \sum_{n=0}^{n=S} [S!/((S-n)!n!)]\ \Delta^{S-n} Y_0 \times X_n$

Solving Large-Scale Linear Systems

- Original problem: d/dt $\mathbf{x}_t = \mathbf{A}\,\mathbf{x}_t$
- 1. Map problem to new space: $\Delta\,\mathbf{X}_S = \mathbf{A}\,\mathbf{X}_S$
- 2. Solve problem in this new space:

  $$\mathbf{X}_{S+1} = (\mathbf{1} + \mathbf{A})\,\mathbf{X}_S \Rightarrow \mathbf{X}_S = (\mathbf{1} + \mathbf{A})^S\,\mathbf{X}_0$$
- 3. Map solution back to original space: $\mathbf{x}_t = e^{\mathbf{A}t}\,\mathbf{x}_0$
  - *And you are done*

- Some applications:
  - *Master equations and random walks*
  - *Model-based optimal control theory*

Non-linear dynamics (and chaos)

- Example: So-called logistic models
  - ➔ E.g. found in ecology, and diffusion/learning theory
  - ➔ Note: Discrete models yield basic example of chaos
- ODE equation: $d/dt\, x_t = A\, x_t\, (1 - x_t)$
- Discrete analogy $\Delta X_{S+1} = A\, X_S\, (1 - X_S)$ **? No!**

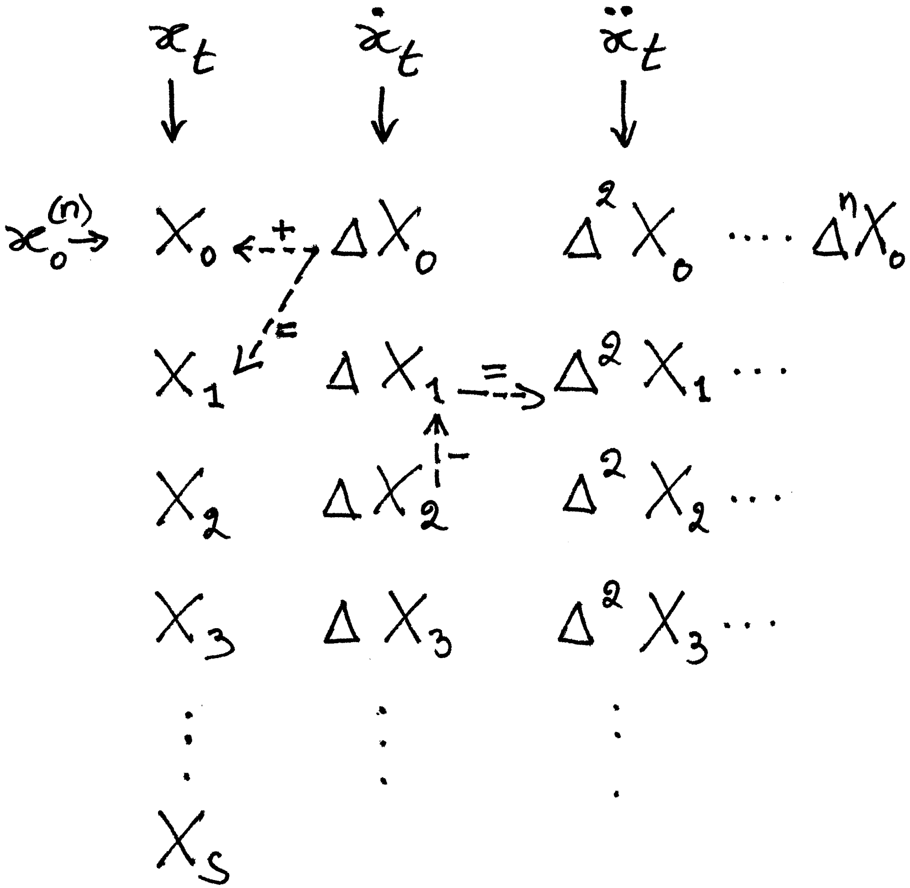
$$X_{S+1} = (1+A) X_S - A \sum_{n=0}^{S} \binom{S}{n} \Delta^{S-n} X_0 \cdot X_n$$

Similarly for famous Lorenz (“butterfly”) chaos [weather]

- Unexpected effects:
  - ➔ No memory in continuous system
  - ➔ But memory in equivalent discrete system

Qualitative Physics / Reasoning

- Shoham's Extended Prediction Problem Does Not Exist
- Strictly discrete and finite reasoning can give you all the results of infinitesimal calculus
- See T Transform tableau method



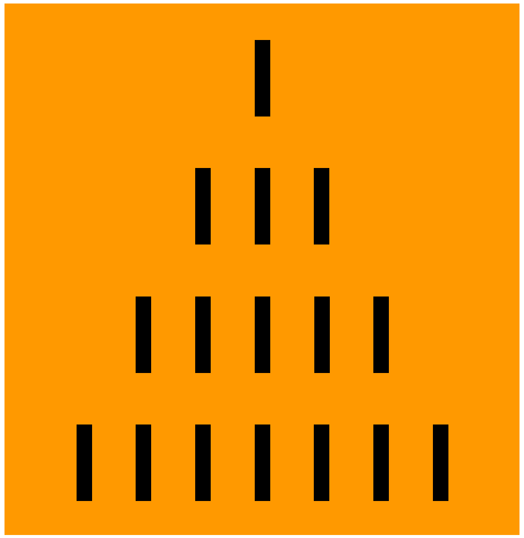
Conclusion: Android Epistemology

- It is indeed possible to solve continuous problems in strictly discrete ways by ontology mappings
- Thus, discrete beings (such as androids) have the same capabilities as continuous beings (such as humanoids)

- Ontology mappings are themselves sophisticated ontologies, not simply (equivalence) concept maps
- Ontology mapping is a substantive issue, rather than a matter of formal representation
  - ➔ This already applies to upper level ontologies such as time
  - ➔ (so, what about the rest?)
- There is a need for substantive ontology modelling (not representation) work and applications
  - ➔ There is not enough of this in the Semantic Web field now

Appendix: Nim

- Rules (two players):
  - ➔ Each turn: delete/remove any desired number of sticks (stones), but from one row (box) *only*
  - ➔ Who is left with the *last* stick (stone), looses the game



Towards Browsing Distant Metadata Using Semantic Signatures

Andrew Choi & Marek Hatala

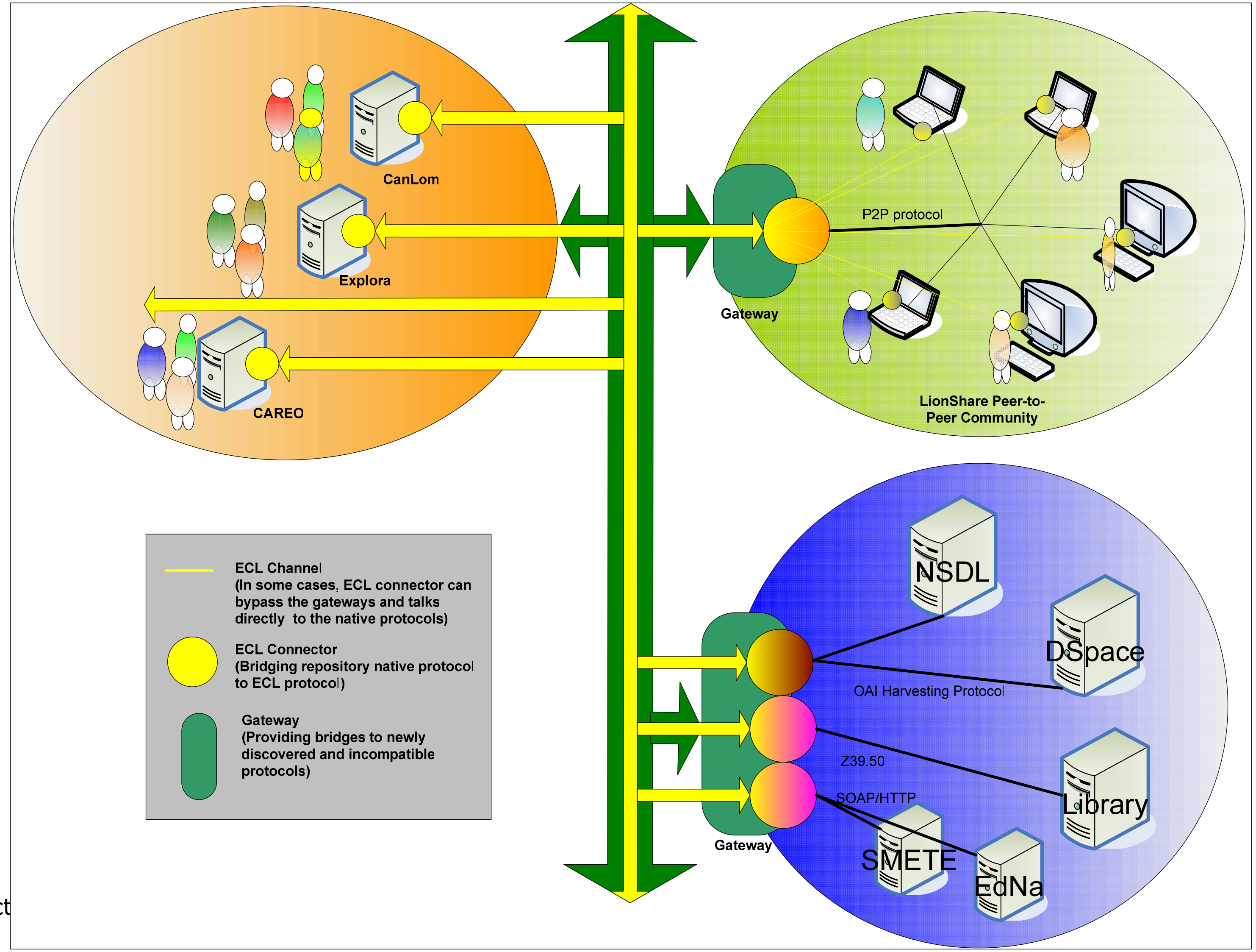
Laboratory for Ontological Research
School of Interactive Arts and Technology
Simon Fraser University Surrey
http://lore.iat.sfu.ca

K-CAP Workshop on Integrating Ontologies
October 2, 2005 in Banff, Canada

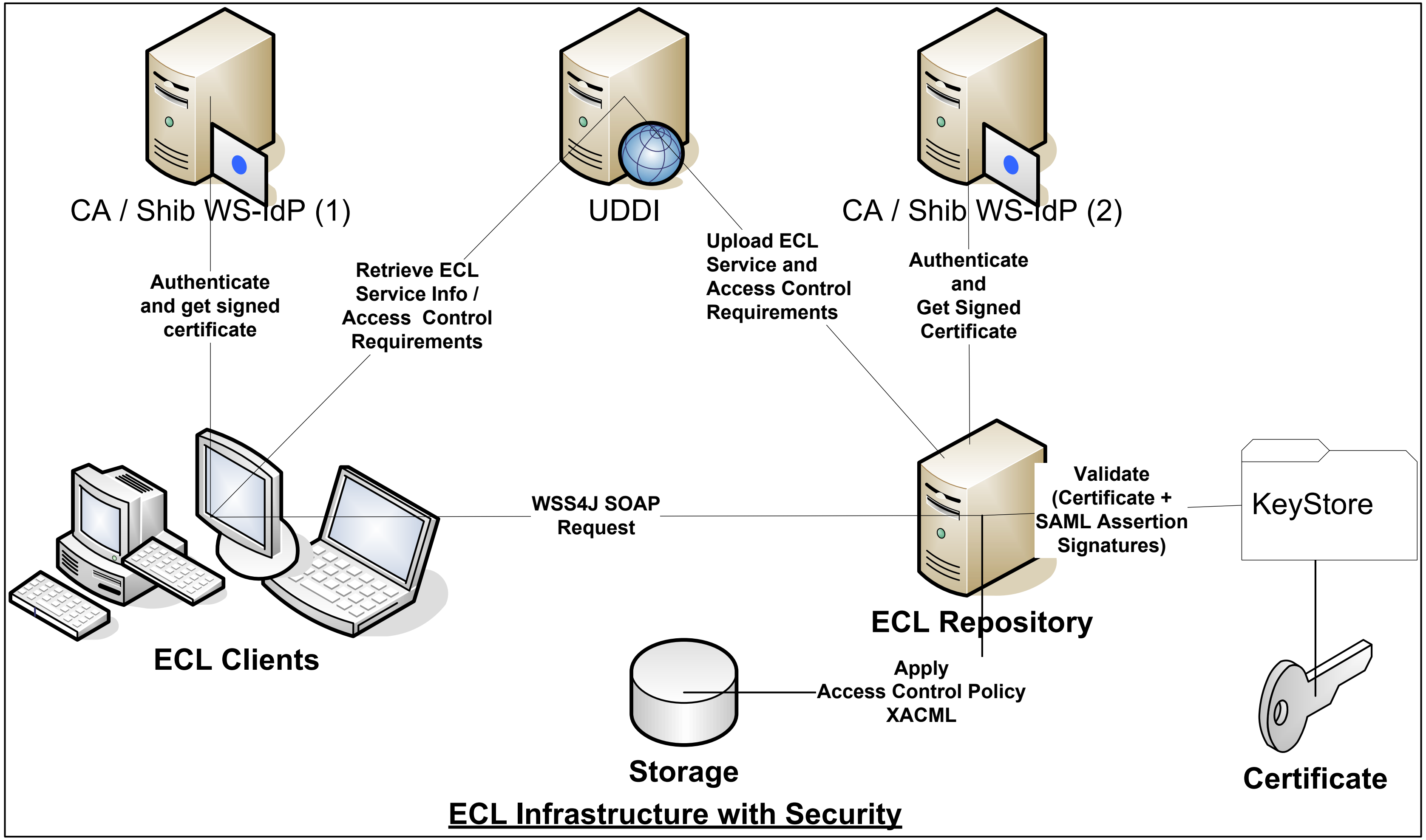
Overview

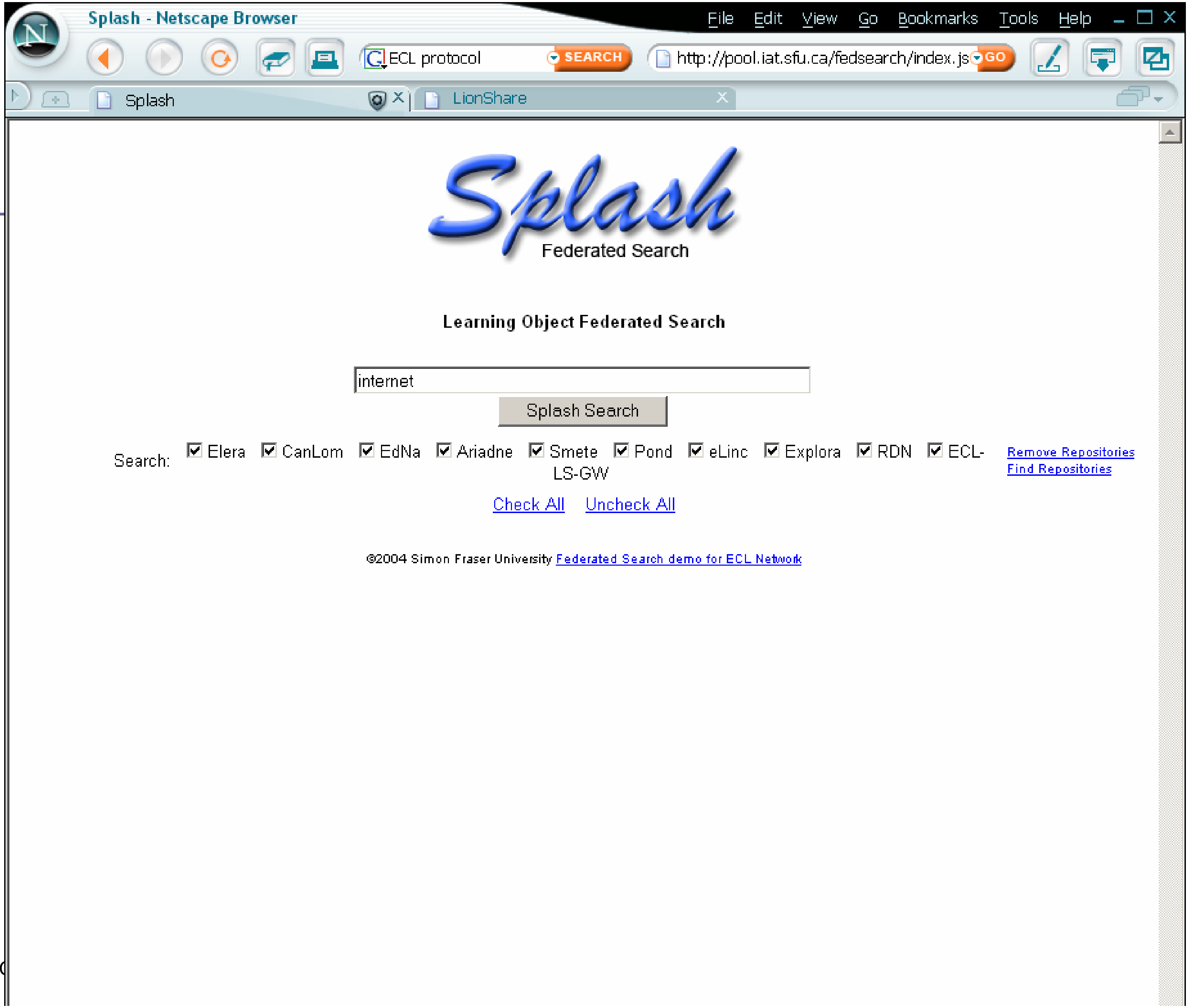
- Background and Motivation
- Approach
- Realization
- Evaluation & Discussion
- Future Directions

Open Network of Learning Services

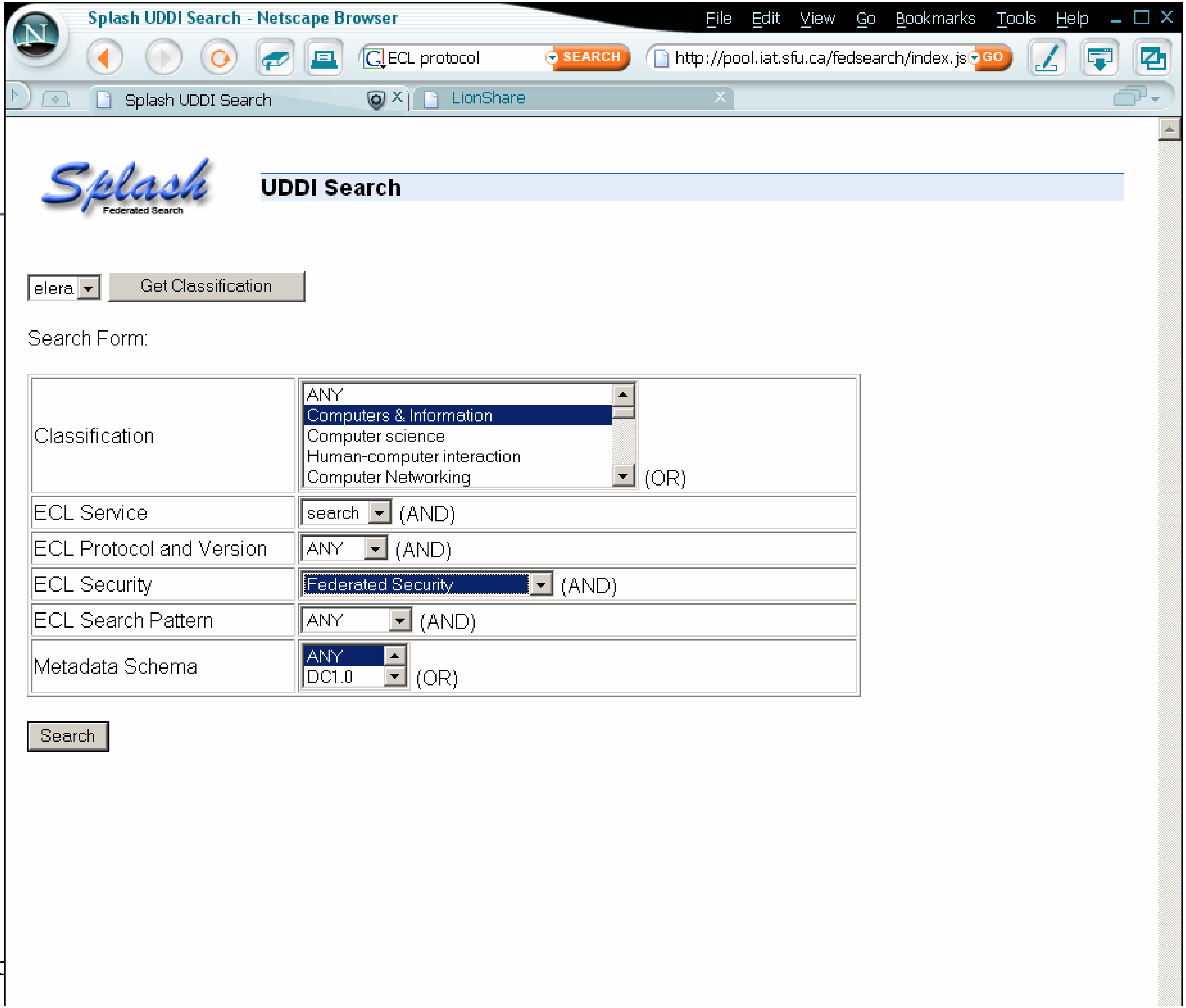


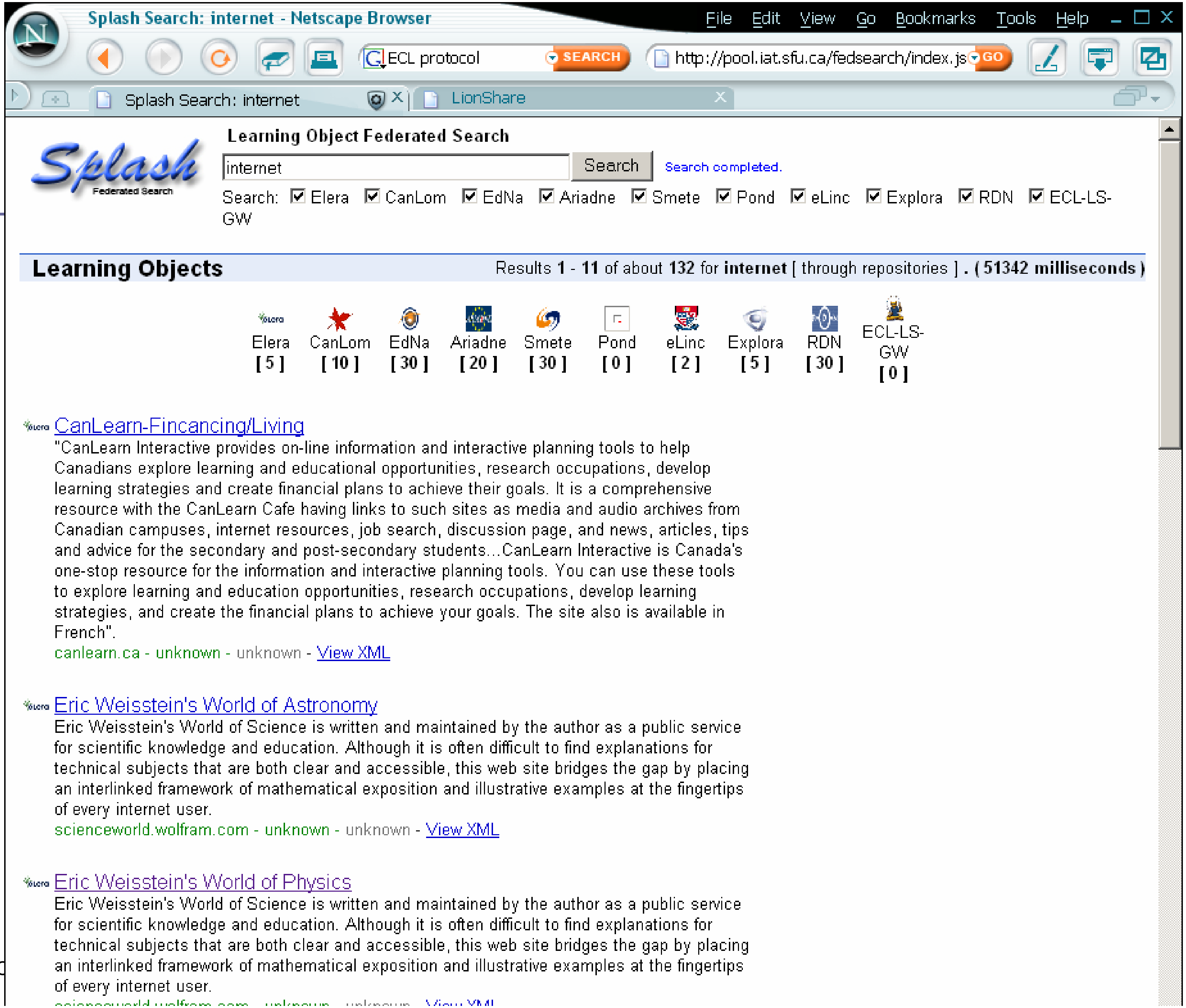
ECL Infrastructure





Splash - Netscape Browser
File Edit View Go Bookmarks Tools Help
ECL protocol
SEARCH
http://pool.iat.sfu.ca/fedsearch/index.js
GO
Splash
LionShare
Splash
Federated Search
Learning Object Federated Search
internet
Splash Search
Search: Elera CanLom EdNa Ariadne Smete Pond eLinc Explora RDN ECL-LS-GW
Remove Repositories
Find Repositories
Check All Uncheck All
©2004 Simon Fraser University Federated Search demo for ECL Network


Splash UDDI Search - Netscape Browser
File Edit View Go Bookmarks Tools Help
ECL protocol
SEARCH
http://pool.iat.sfu.ca/fedsearch/index.js
GO
Splash UDDI Search
LionShare
Splash
Federated Search
UDDI Search
elera
Get Classification
Search Form:
Classification
ANY
Computers & Information
Computer science
Human-computer interaction
Computer Networking
(OR)
ECL Service
search
(AND)
ECL Protocol and Version
ANY
(AND)
ECL Security
Federated Security
(AND)
ECL Search Pattern
ANY
(AND)
Metadata Schema
ANY
DC1.0
(OR)
Search



Splash Search: internet - Netscape Browser
File Edit View Go Bookmarks Tools Help
ECL protocol
SEARCH
http://pool.iat.sfu.ca/fedsearch/index.js
GO
Splash Search: internet
LionShare
Splash
Federated Search
Learning Object Federated Search
internet
Search
Search completed.
Search: Elera CanLom EdNa Ariadne Smete Pond eLinc Explora RDN ECL-LS-GW
Learning Objects
Results 1 - 11 of about 132 for internet [through repositories] . (51342 milliseconds)
Elera [5]
CanLom [10]
EdNa [30]
Ariadne [20]
Smete [30]
Pond [0]
eLinc [2]
Explora [5]
RDN [30]
ECL-LS-GW [0]
CanLearn-Fincancing/Living
"CanLearn Interactive provides on-line information and interactive planning tools to help Canadians explore learning and educational opportunities, research occupations, develop learning strategies and create financial plans to achieve their goals. It is a comprehensive resource with the CanLearn Cafe having links to such sites as media and audio archives from Canadian campuses, internet resources, job search, discussion page, and news, articles, tips and advice for the secondary and post-secondary students...CanLearn Interactive is Canada's one-stop resource for the information and interactive planning tools. You can use these tools to explore learning and education opportunities, research occupations, develop learning strategies, and create the financial plans to achieve your goals. The site also is available in French".
canlearn.ca - unknown - unknown - View XML
Eric Weisstein's World of Astronomy
Eric Weisstein's World of Science is written and maintained by the author as a public service for scientific knowledge and education. Although it is often difficult to find explanations for technical subjects that are both clear and accessible, this web site bridges the gap by placing an interlinked framework of mathematical exposition and illustrative examples at the fingertips of every internet user.
scienceworld.wolfram.com - unknown - unknown - View XML
Eric Weisstein's World of Physics
Eric Weisstein's World of Science is written and maintained by the author as a public service for scientific knowledge and education. Although it is often difficult to find explanations for technical subjects that are both clear and accessible, this web site bridges the gap by placing an interlinked framework of mathematical exposition and illustrative examples at the fingertips of every internet user.

Motivation Summary

- Distributed network of object repositories
- Users select repositories as they become available
- No prior alignment of conceptual structures between repositories

- Goal: Support search and retrieval using local concepts

Approach: Semantic Signatures

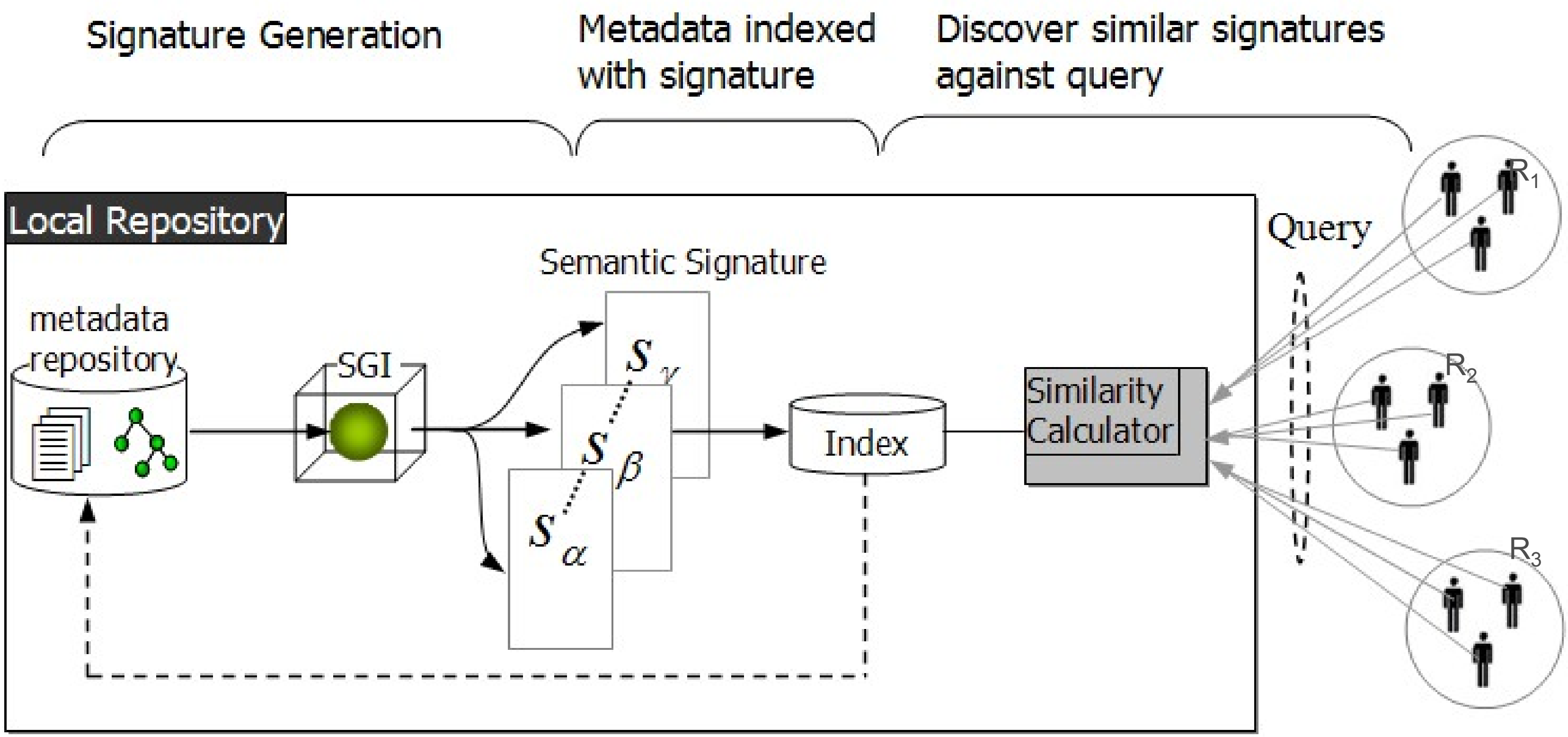
Assumptions

- Content: ***metadata*** (+ objects)
- Content in the remote repository annotated with ***remote*** concepts
- User ***associated*** with local content (user community repository, individual collection, etc.)
- Local content annotated with ***local*** concepts

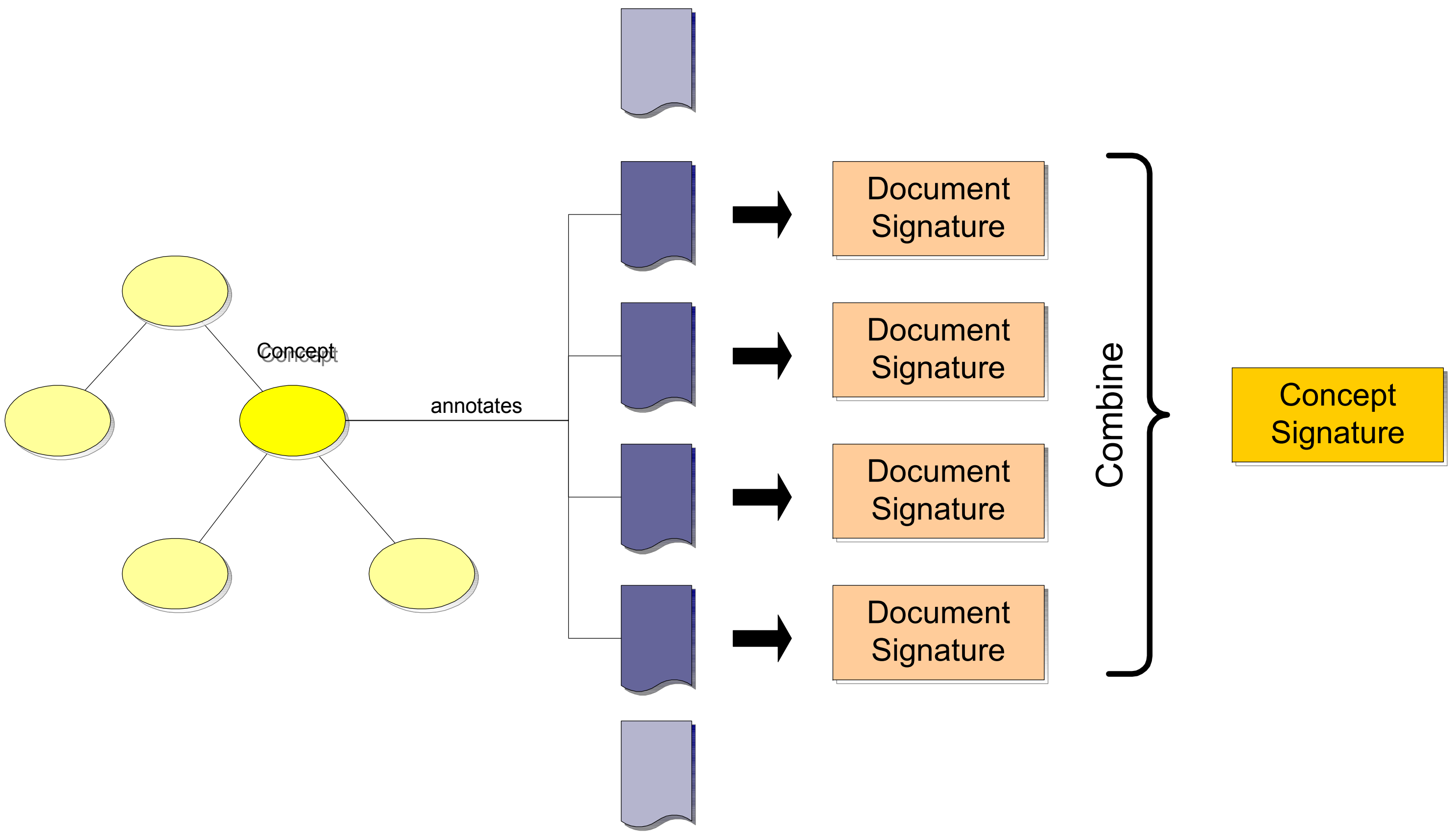
Main Idea

- Use WordNet as a mediator
- Represent concepts in the ontology using semantic signatures
  - *Semantic signature is a logical grouping of representational word senses for the concept.*
- Match signatures to determine concept similarity between local and remote concepts

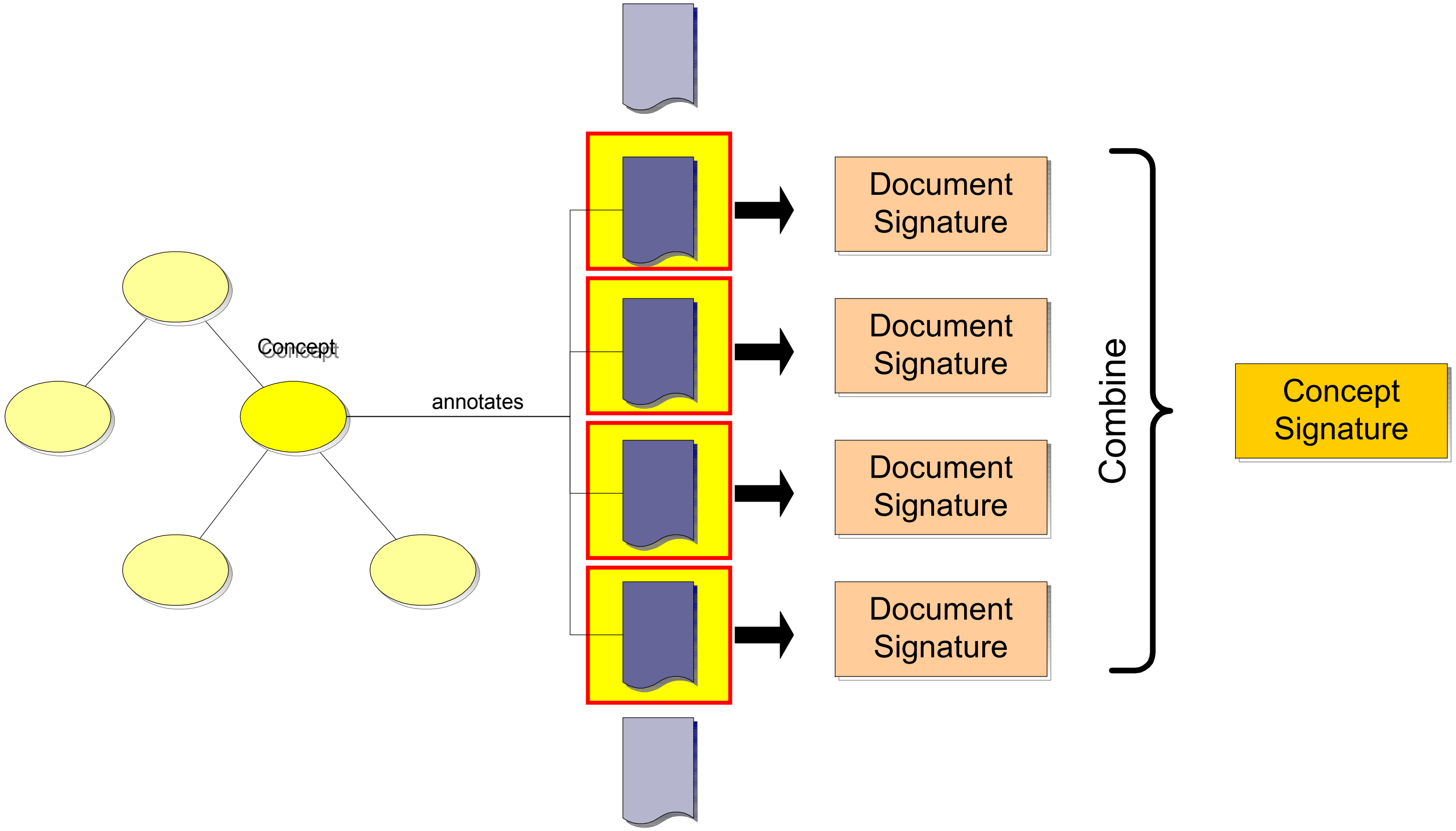
Searching with Signatures



Signature Generation Steps

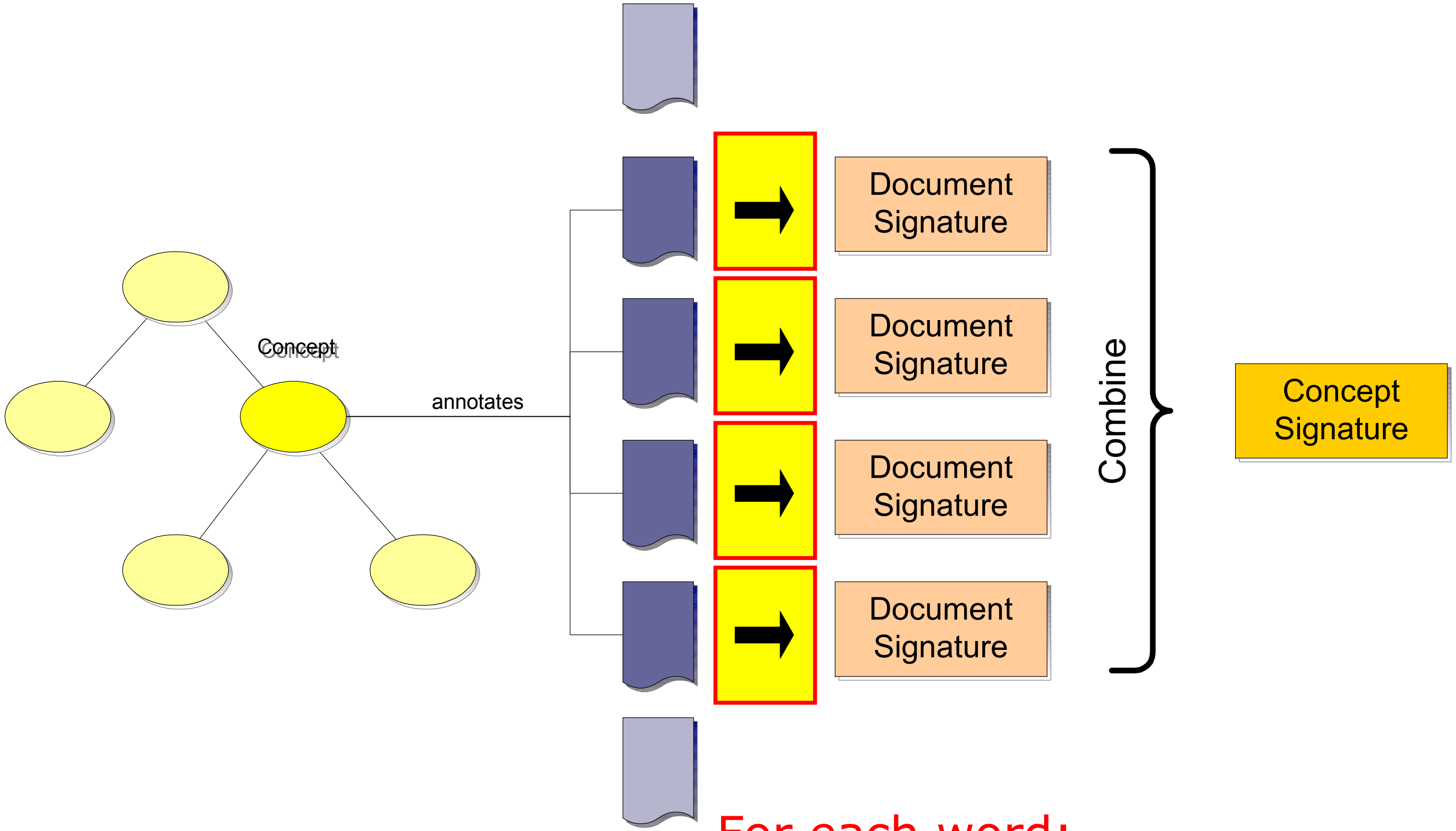


Signature Generation Steps



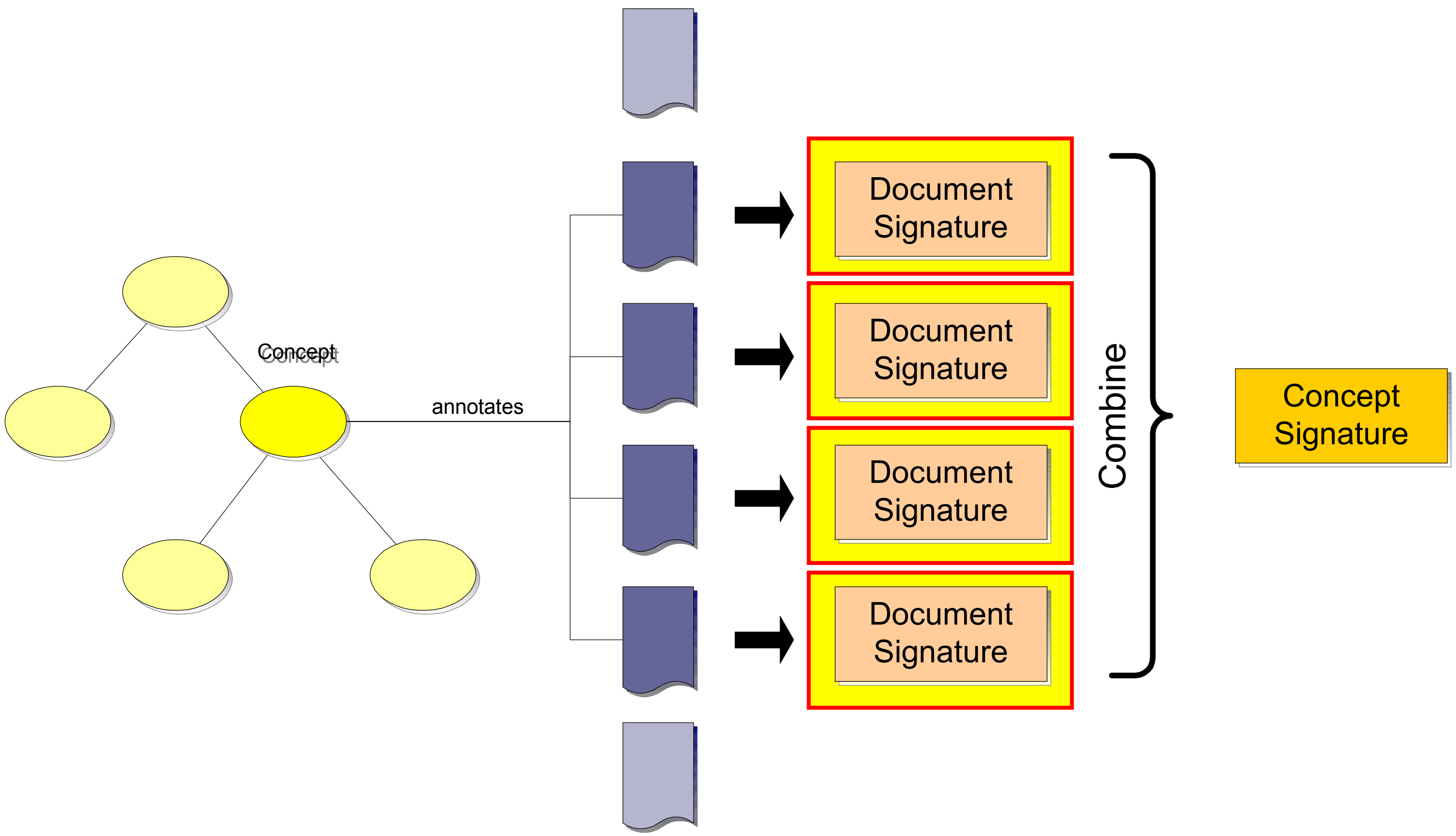
TFIDF across all documents annotated with same concept

Signature Generation Steps



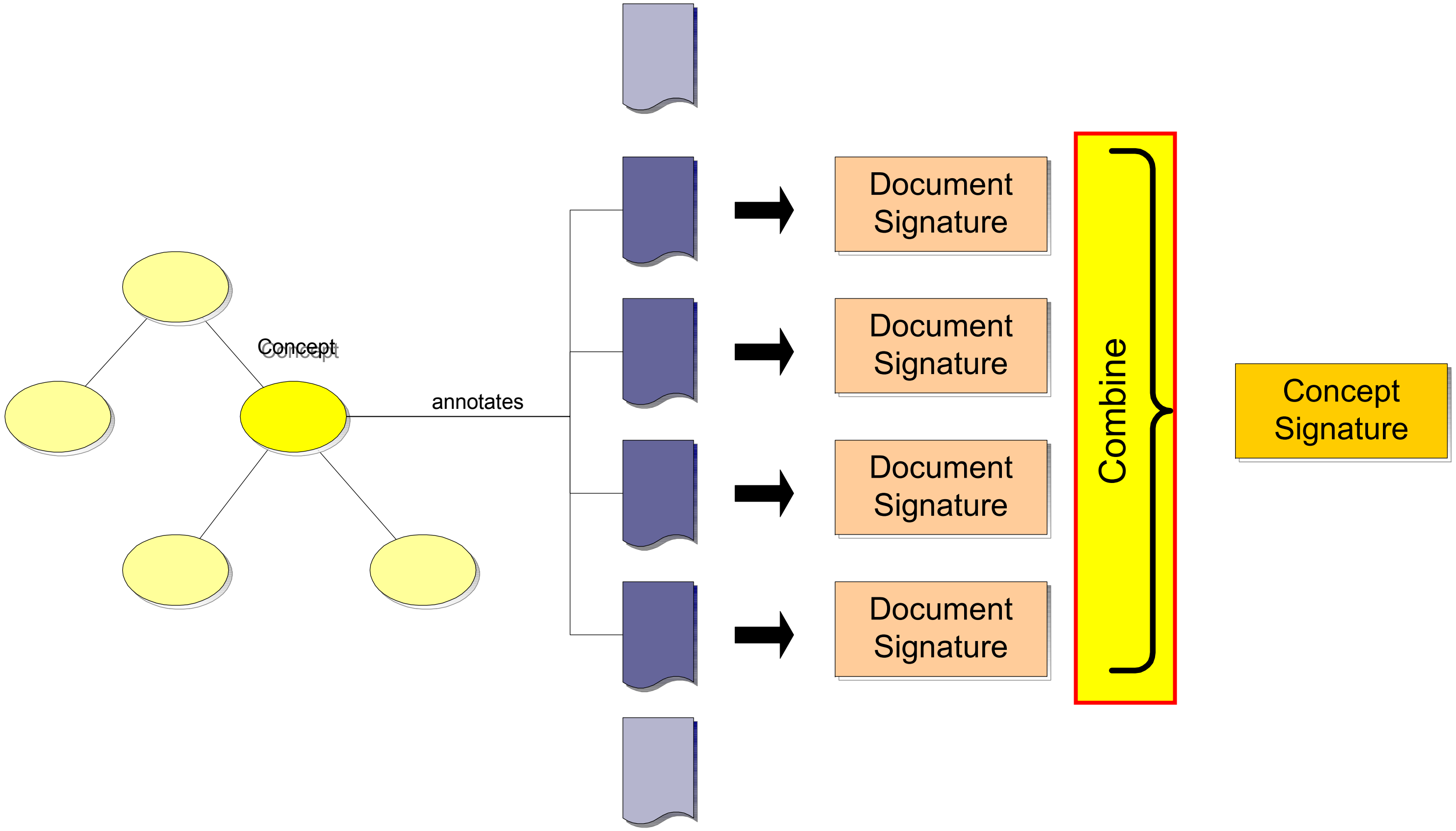
For each word:
Retrieve senses from Wordnet,
Best Sense Selection

Signature Generation Steps



Signatures as sets of senses

Signature Generation Steps



TFIDF across senses in document signatures

Best Sense Selection

For each metadata document $D \in C_1$

Get the list of synsets for each word term $T_1 \in D$

For each synset Syn_1 of the word term T_1

For each sense term $S_i \in Syn_1$

➡ 1 Compute associative frequency af for S_i to other senses $S_k \in Syn_k$, $Syn_k \subseteq T_k$ and $T_1 \neq T_k$

1.1 return the sense S_l with highest score Max(af)

2 Compute associative frequency af for S_i to k-order parent senses $PS_k \in P(Syn_k)$, $P(Syn_k) \subseteq T_k$ and $T_1 \neq T_k$

2.1 return the sense S_p with highest score Max(af)

3 Record the most popular sense S_w offered by WordNet

Select the sense according to the preference ranking to represent the word term T_1

Return the Best Sense to represent word term T_1

Aggregate all sense from all important word terms to represent signature of the document D

Strategy I: Local Context

- Example: Windows is an OS for computer system.

Document vector D_1

Word term	Word Sense
Windows	Synset 1: ⟨windowpane, window⟩ Synset 2: ⟨operating system, computer screen⟩ Synset 3: ⟨framework, opening⟩
OS	Synset 1: ⟨os⟩ Synset 2: ⟨osmium, Os, atomic number 76⟩ Synset 3: ⟨operating system, OS⟩ Synset 4: ⟨oculus sinister, OS⟩
Computer	Synset 1: ⟨computer device, computing machine, data processor⟩ Synset 2: ⟨calculator, reckoner, figurer, estimator, computer⟩

- "Synset 2" will be selected as the best sense for word "Windows"

Strategy II: Parent Senses

For each metadata document $D \in C_1$

Get the list of synsets for each word term $T_1 \in D$

For each synset Syn_1 of the word term T_1

For each sense term $S_i \in Syn_1$

1. Compute associative frequency af for S_i to other senses $S_k \in Syn_k$, $Syn_k \subseteq T_k$ and $T_1 \neq T_k$

   1.1 return the sense S_l with highest score Max(af)

2. Compute associative frequency af for S_i to k-order parent senses $PS_k \in P(Syn_k)$, $P(Syn_k) \subseteq T_k$ and $T_1 \neq T_k$

   2.1 return the sense S_p with highest score Max(af)

3. Record the most popular sense S_w offered by WordNet

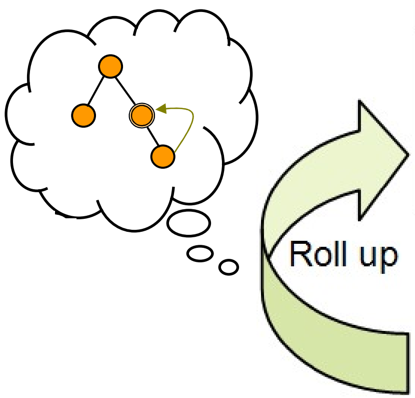
Select the sense according to the preference ranking to represent the word term T_1

Return the Best Sense to represent word term T_1

Aggregate all sense from all important word terms to represent signature of the document D

Strategy II: Most Specific Parent

- Example: Sun is the center of our <u>solar system</u>.



Hypernyms of	Word sense
Sun	(P)Synset 1: ⟨star⟩
	(P)Synset 2: ⟨light, visible light, radiation⟩
	⋮

Word term	Word sense
Sun	Synset 1: ⟨sunlight, sunshine, sun⟩ Synset 2: ⟨Sunday, Lord's Day, Dominicus, Sun⟩ Synset 3: ⟨sun⟩
Solar system	Synset 1: ⟨star, system, star scheme⟩

- "*(P)Synset 1*" will be selected as the best sense for word "*Sun*"

Strategy III: Frequency

For each metadata document $D \in C_1$

Get the list of synsets for each word term $T_1 \in D$

For each synset Syn_1 of the word term T_1

For each sense term $S_i \in Syn_1$

1. Compute associative frequency af for S_i to other senses $S_k \in Syn_k$, $Syn_k \subseteq T_k$ and $T_1 \neq T_k$

   1.1 return the sense S_1 with highest score Max(af)

2. Compute associative frequency af for S_i to k-order parent senses $PS_k \in P(Syn_k)$, $P(Syn_k) \subseteq T_k$ and $T_1 \neq T_k$

   2.1 return the sense S_p with highest score Max(af)

3. Record the most popular sense S_w offered by WordNet

Select the sense according to the preference ranking to represent the word term T_1

Return the Best Sense to represent word term T_1

Aggregate all sense from all important word terms to represent signature of the document D

Evaluation Experiment

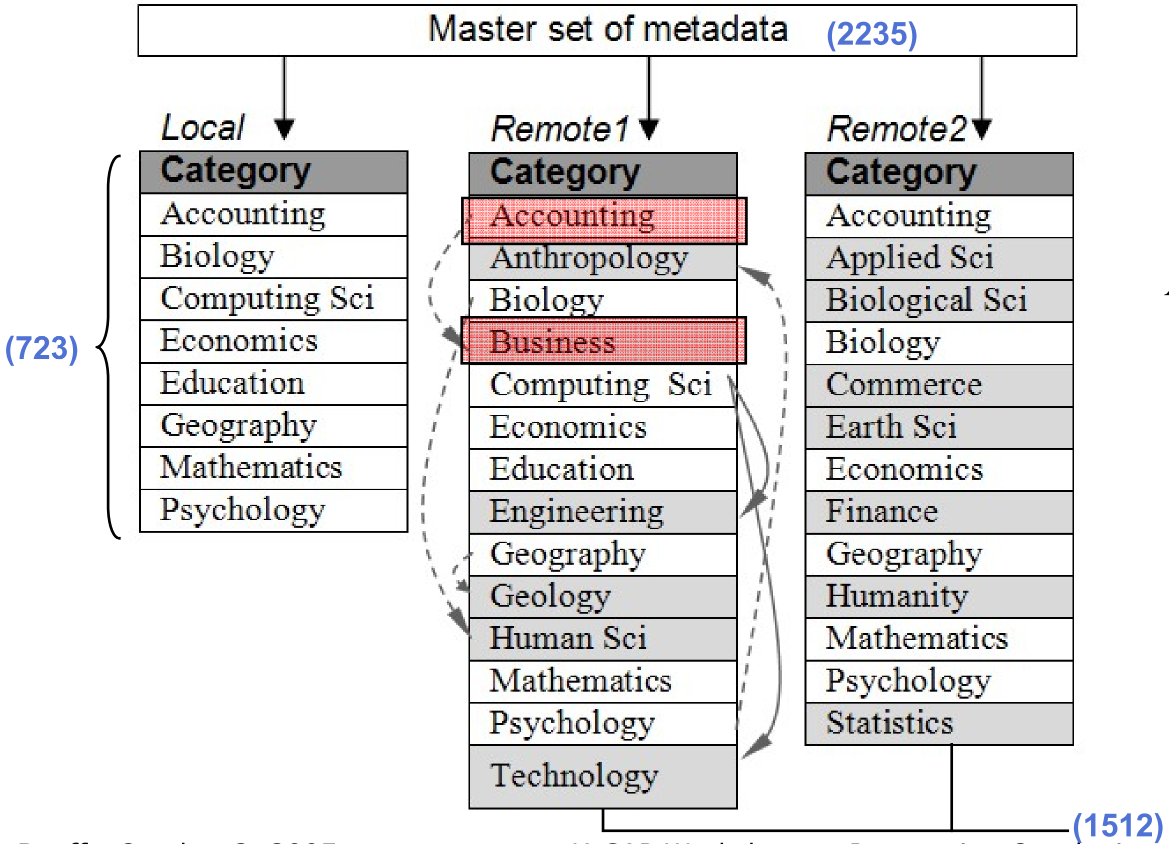
- 3 independent databases are set up: ***local***, ***remote1*** and ***remote2***
- *Local* represents the local repository (training dataset)
- *remote1* and *remote2* represent distant repositories (testing dataset)
- Effectiveness of retrieval measured by number of relevant concepts returned from remote repositories

Dataset

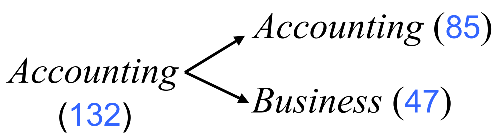
- 8 different categories of 2235 metadata are acquired from various sources

Category	Sources	No. of metadata
Accounting	Business Source Premier Publications	382
Biology	Biological and Agricultural Index, BioMed Central Online Journals	315
Computing Science	Citeseer	320
Economics	American Economic Association's electronic database	353
Education	Educational Resource Information Center	307
Geography	Geobase	237
Mathematics	arXiv.org, MathSciNet	157
Psychology	PsycINFO, ERIC	164
	Total	**2235**

- Metadata are distributed randomly to training and testing group



For example:



Results

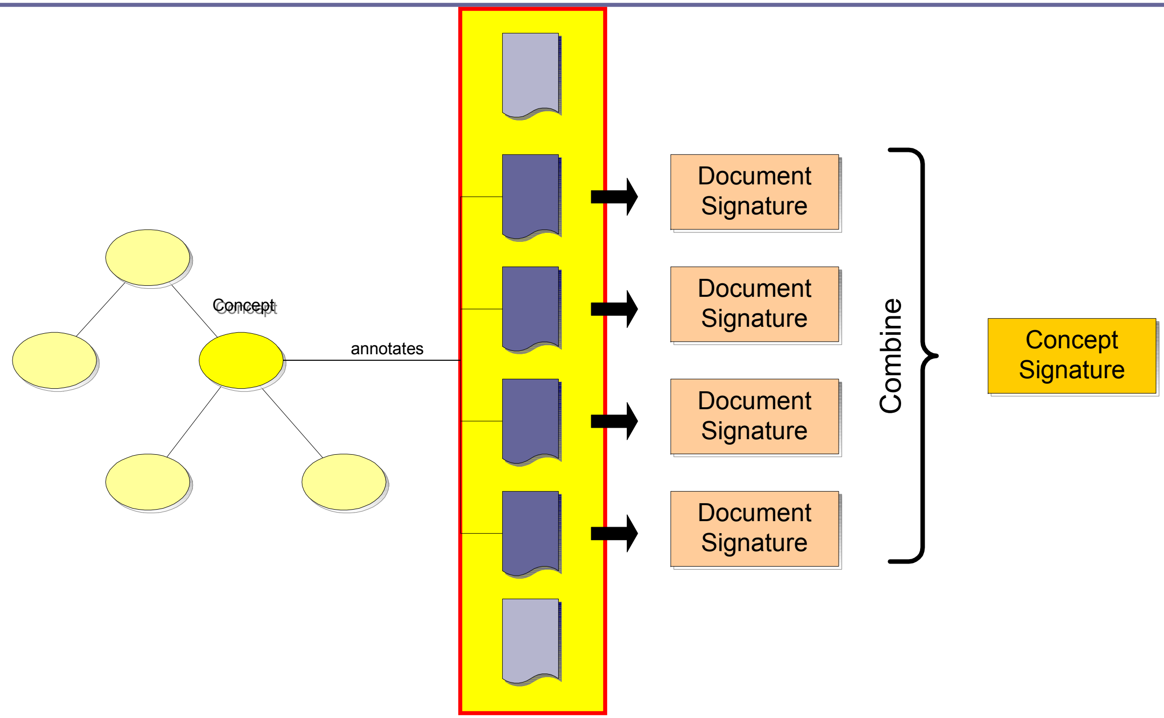
Category	Precision		Recall		F-measure	
	S	*K*	*S*	*K*	*S*	*K*
Accounting	1.00	0.67	1.00	0.75	1.00	0.71
Biology	0.75	0.75	0.75	0.75	0.75	0.75
Computing Sci	1.00	0.50	1.00	0.50	1.00	0.50
Economic	1.00	0.75	1.00	0.75	1.00	0.86
Education	1.00	0.50	1.00	0.75	1.00	0.45
Geography	0.75	0.50	0.75	0.50	0.75	0.50
Mathematics	0.67	0.33	0.67	0.50	0.67	0.40
Psychology	0.67	0.33	0.67	0.67	0.67	0.44
Average	0.86	0.54	0.86	0.65	0.86	0.58

S = Signature-based
K = Keywords-based

Discussion

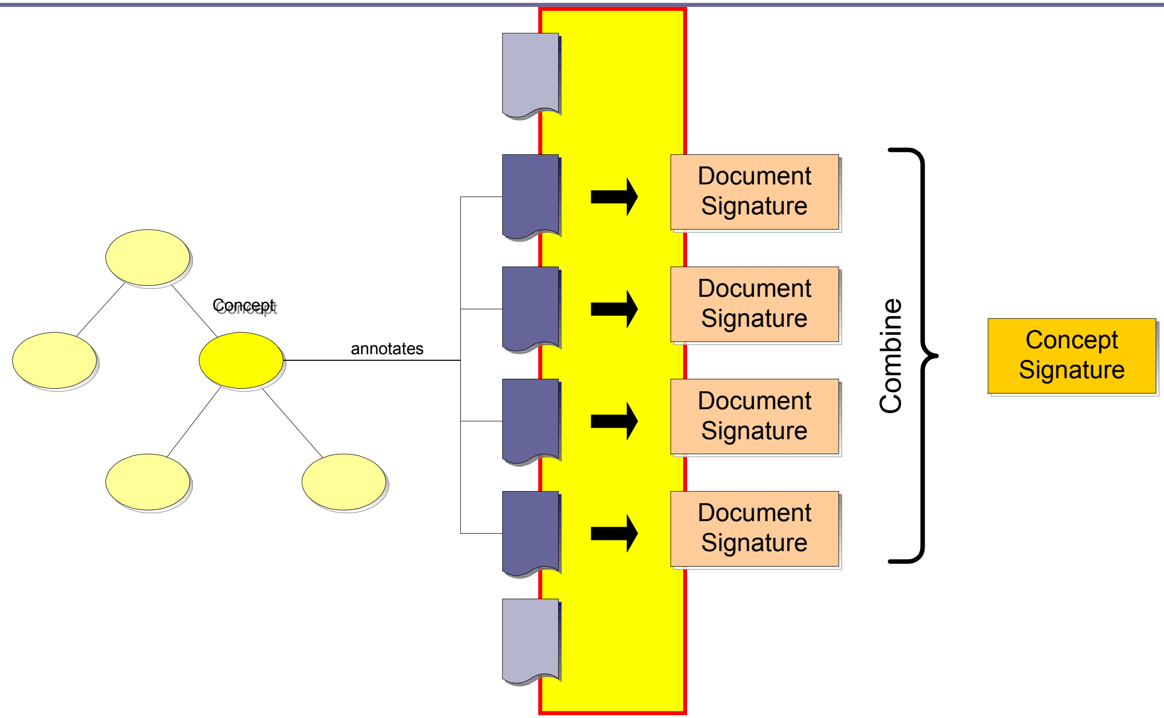
- Only basic evaluation but shows promise
- Many unanswered questions:
  - Number/Size of the documents per concept
  - Nature of the documents
  - Specificity of the domain
  - Shifting context

Future: Text Processing



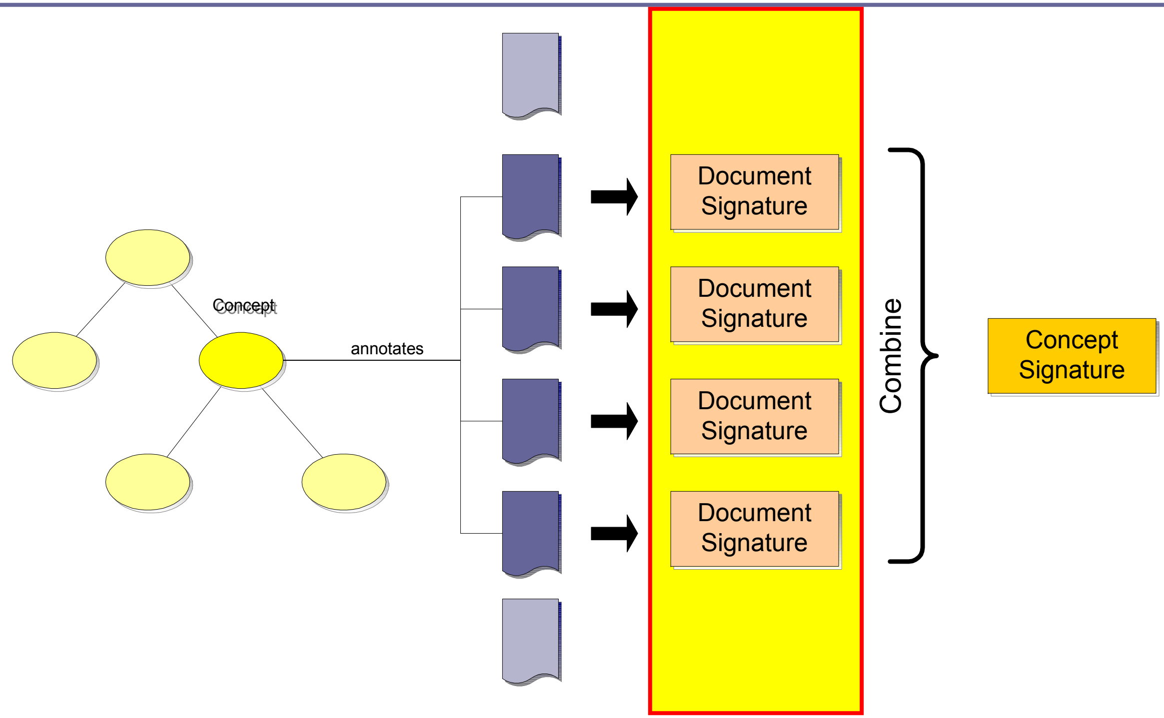
- summarization, significant phrases, etc.

Future: Sense Disambiguation



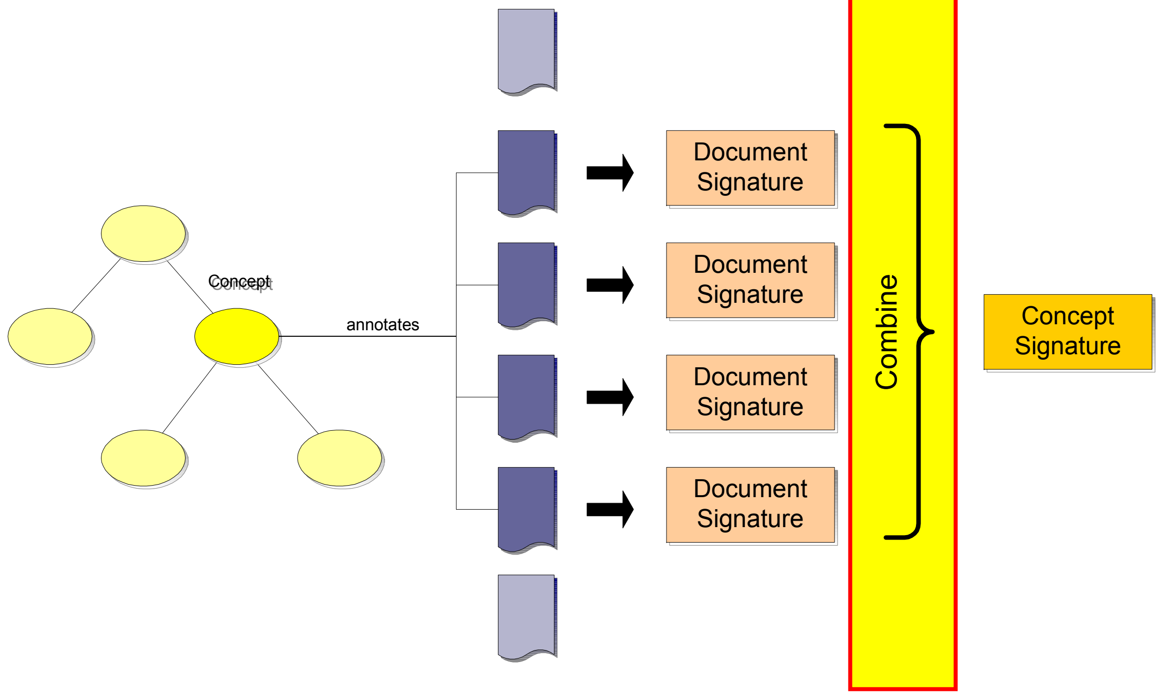
- explore WordNet structure and document structure (now only direct parent)

Future: Signatures



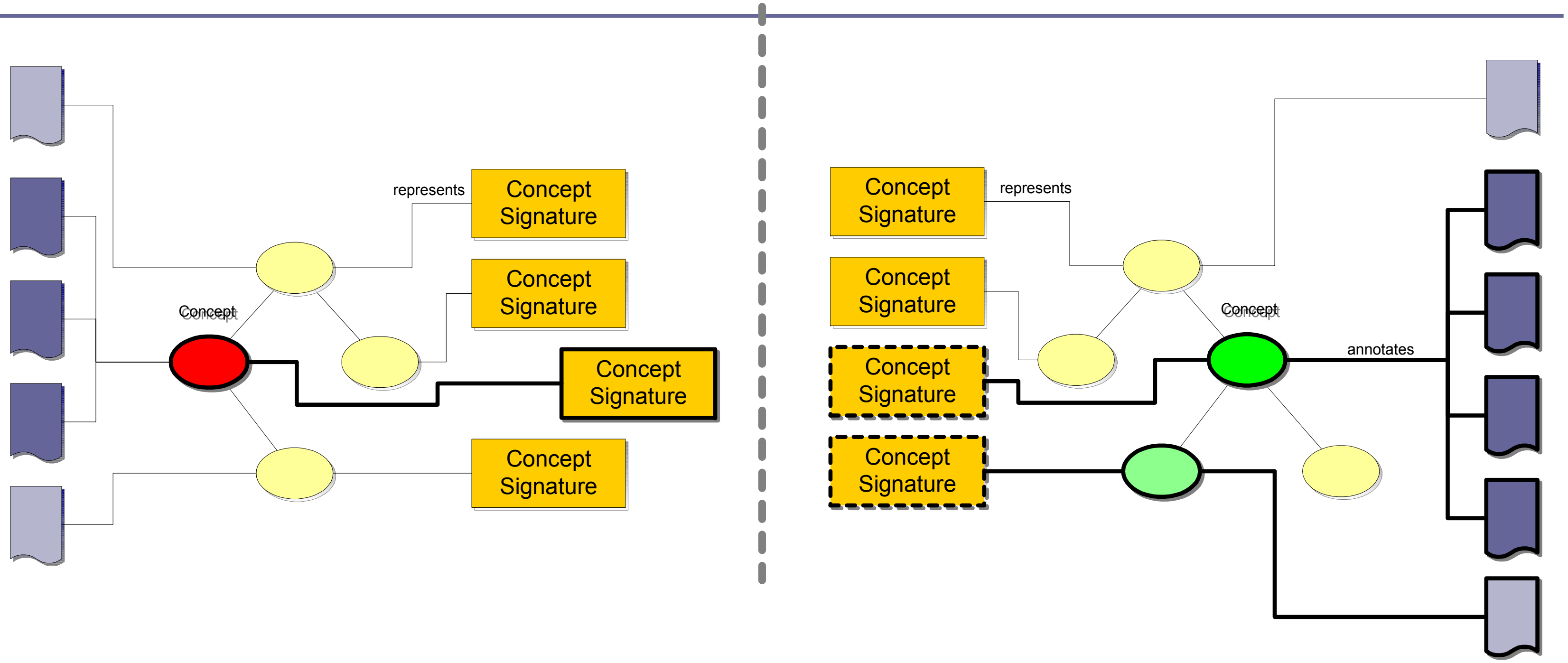
- structure, include domain-specific words not found in WordNet, etc.
- Concept differentiation capability

Future: Signature Combination



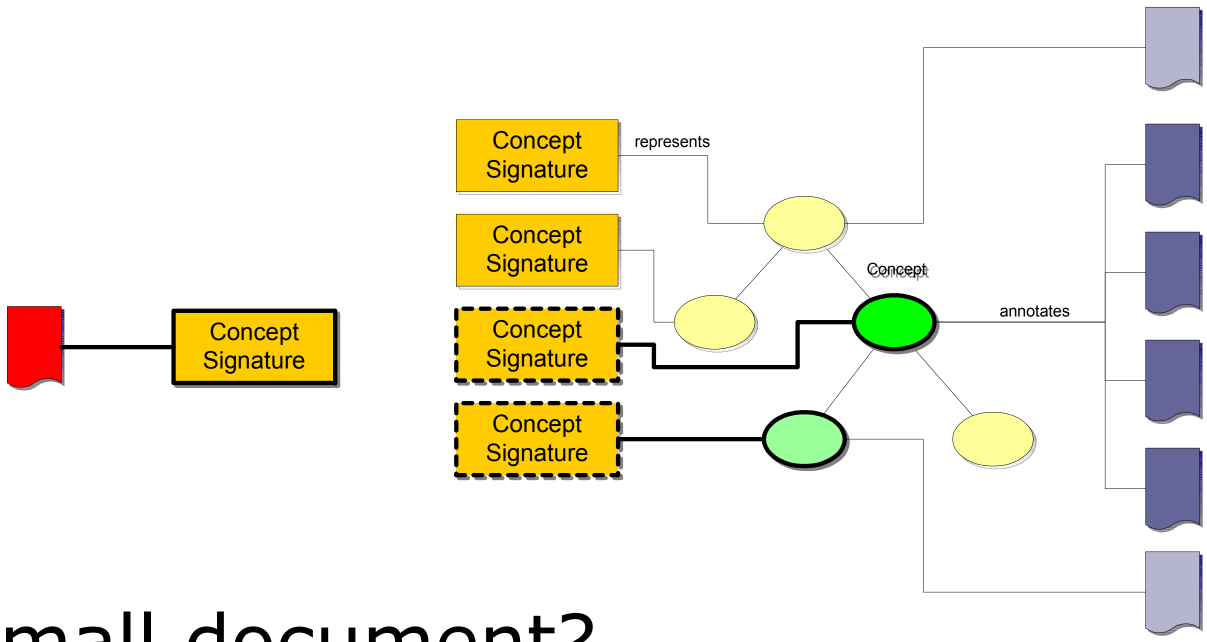
- Utilize WordNet structure when merging signatures
- Utilize ontology structure (?)

App: Distributed Search with Concepts



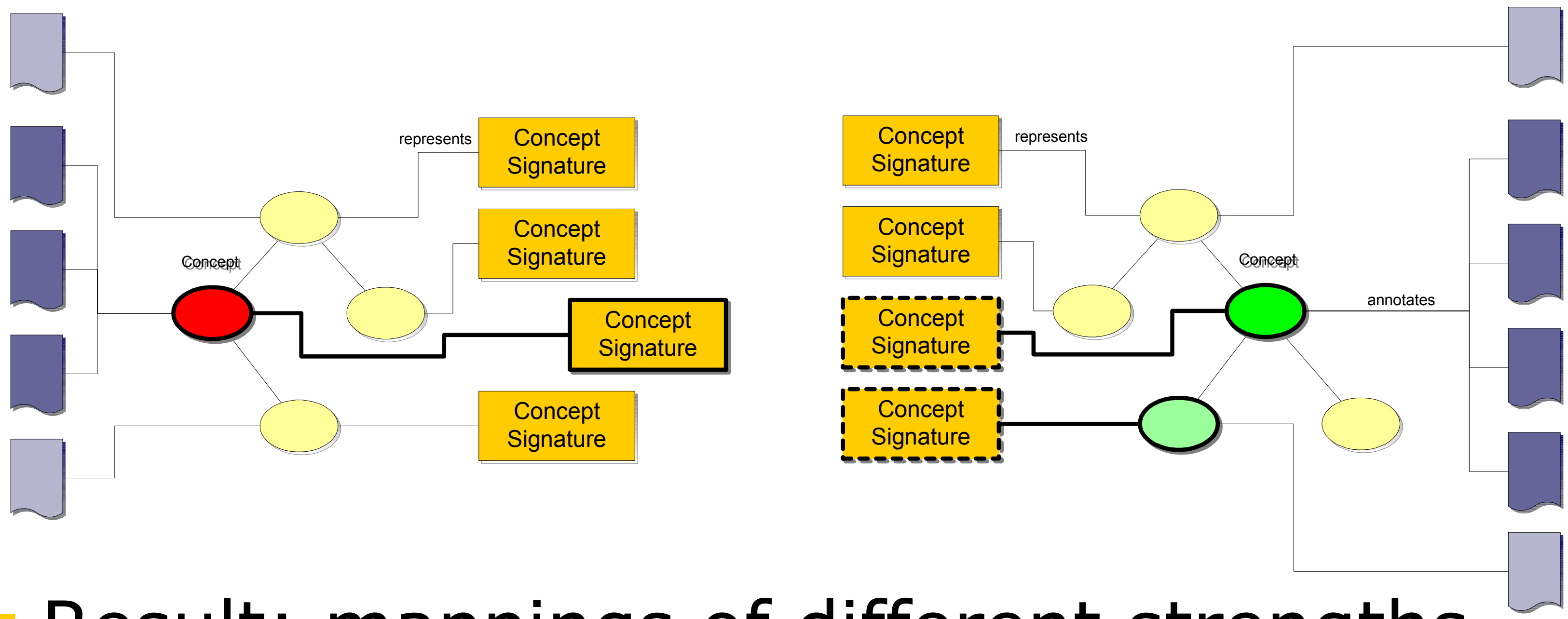
- Current implementation
- Support in the middleware

App: New Document Annotation



- How big/small document?
- Thresholds issue
- Signature libraries for well known classifications (ACM CCS) ➔ web-service

App: Ontology Alignment



- Result: mappings of different strengths
- Threshold issue
- As a complement to other methods

Comments and Suggestions

- Email: mhatala@sfu.ca

Semantic Association of Taxonomy-based Standards Using Ontology

Hung-Ju Chu, Randy Y. C. Chow, Su-Shing Chen

Computer and Information Science and Engineering

Raja R.A. Issa, Ivan Mutis

Rinker School of Building Construction

University of Florida

10/2/2005

Research Focus

- Methodology for matching complementary taxonomies (hierarchically structured standards) to facilitate cross-referencing required in workflows.

Target Application

- Building Construction Domain
- Masterformat [1] &UniformatII [2]

First level of Masterformat

Division Numbers and Titles

Procurement and Contracting Requirements Group

Division 00 Procurement and Contracting Requirements

Specifications Group

General Requirements Subgroup

Division 01 General Requirements

Facility Construction Subgroup

Division 02 Existing Conditions
Division 03 Concrete
Division 04 Masonry
Division 05 Metals
Division 06 Wood, Plastics, and Composites
Division 07 Thermal and Moisture Protection
Division 08 Openings
Division 09 Finishes
Division 10 Specialties
Division 11 Equipment
Division 12 Furnishings
Division 13 Special Construction
Division 14 Conveying Equipment
Division 15 Reserved
Division 16 Reserved
Division 17 Reserved
Division 18 Reserved
Division 19 Reserved

Site and Infrastructure Subgroup

Division 30 Reserved
Division 31 Earthwork
Division 32 Exterior Improvements
Division 33 Utilities
Division 34 Transportation
Division 35 Waterway and Marine Construction
Division 36 Reserved
Division 37 Reserved
Division 38 Reserved
Division 39 Reserved

Process Equipment Subgroup

Division 40 Process Integration
Division 41 Material Processing and Handling Equipment
Division 42 Process Heating, Cooling, and Drying Equipment
Division 43 Process Gas and Liquid Handling, Purification, and Storage Equipment
Division 44 Pollution Control Equipment

First 3 levels of UniformatII

ASTM Uniformat II Classification for Building Elements (E1557-97)

Level 1 Major Group Elements	Level 2 Group Elements	Level 3 Individual Elements
A SUBSTRUCTURE	A10 Foundations	A1010 Standard Foundations
		A1020 Special Foundations
		A1030 Slab on Grade
	A20 Basement Construction	A2010 Basement Excavation
		A2020 Basement Walls
B SHELL	B10 Superstructure	B1010 Floor Construction
		B1020 Roof Construction
	B20 Exterior Enclosure	B2010 Exterior Walls
		B2020 Exterior Windows
		B2030 Exterior Doors
	B30 Roofing	B3010 Roof Coverings
		B3020 Roof Openings
C INTERIORS	C10 Interior Construction	C1010 Partitions
		C1020 Interior Doors
		C1030 Fittings
	C20 Stairs	C2010 Stair Construction
		C2020 Stair Finishes
	C30 Interior Finishes	C3010 Wall Finishes
		C3020 Floor Finishes
		C3030 Ceiling Finishes
D SERVICES	D10 Conveying	D1010 Elevators & Lifts
		D1020 Escalators & Moving Walks
		D1090 Other Conveying Systems
	D20 Plumbing	D2010 Plumbing Fixtures
		D2020 Domestic Water Distribution
		D2030 Sanitary Waste
		D2040 Rain Water Drainage
		D2090 Other Plumbing Systems
	D30 HVAC	D3010 Energy Supply
		D3020 Heat Generating Systems
		D3030 Cooling Generating Systems
		D3040 Distribution Systems
		D3050 Terminal & Package Units
		D3060 Controls & Instrumentation
		D3070 Systems Testing & Balancing
		D3090 Other HVAC Systems & Equipment
	D40 Fire Protection	D4010 Sprinklers

Usage of Standards in Workflows

- Cost Estimation
- Code compliance checking

Problems

- The routine workflows are costly.
- $15.8 billion annual interoperability cost in capital facilities industry in 2002

Challenges for Matching

- Objects are classified with complementary views
- Many to many matching
- Mapping semantics are implicit
- Taxonomies are changing

Mapping Semantics

B SHELL

B20 EXTERIOR CLOSURE

B2010 EXTERIOR WALLS

- What Masterformat objects should B2010 be mapped to?

Foundation of This Research

- The observation that mapping semantics can be found in project specifications

Project Specification Example

- PPD (Preliminary Project Descriptions)

B SHELL

B20 EXTERIOR CLOSURE

B2010 EXTERIOR WALLS

1. Exterior Wall Framing: Cold-formed, light gage steel studs, C-shape, galvanized finish.

Our Approach

- FORMALIZATION OF TAXONOMY
- ONTOLOGY-BASED SEMANTIC EXTRACTION
- MEASUREMENT OF AFFINITY

FORMALIZATION OF TAXONOMY

- Step 1: relation set identification
- Step 2: relation statements construction
- Step 3: normalization
- Step 4: generalization

Step 1: Relations

- Primitive: unambiguous; static; intrinsic properties of objects; time; space; intention; set relationship
- Derived

Masterformat Example

Division 5- Metals
- 05100 Structural Metal Framing
  - 05120 Structural steel
  - 05140 Structural aluminum
  - 05160 Metal framing systems
- 05400 Cold formed metal framing
  - **05410 Load bearing metal studs**
  - 05420 Cold formed metal joists
  - 05430 Slotted channel framing

Division 6 - Wood and plastics
- 06100 Rough carpentry
  - 06110 Wood framing
- 06400 Architectural woodwork
  - 06460 Wood frames

Relation Examples

- *used_for* (class-class, human intention): purpose
- *kind_of* (class-class, intrinsic): containment relation of attributes of instances.
- *instance_of* (instance-class, intrinsic): membership
- *made_of* (class-class, intrinsic): material component

Table 1. Mathematical Properties of the relations

Relations	Transitive	reflexive	antisymmetric
used_for	-	-	-
kind_of	+	+	+
instance_of	+	+	+
made_of	+	-	-

Step 2: relation statements

- Subject-relation-object triple

Examples:

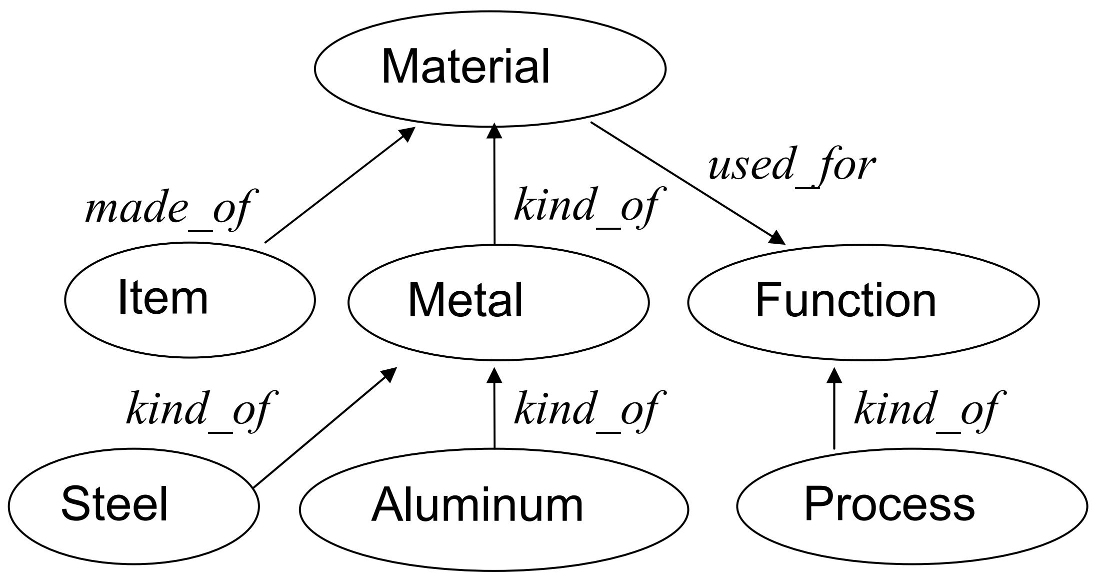
- Metals (D5), Wood (D6), Plastics (D6_1) are *instance_of* Material (root) → (D5_root, D6_root, D6_1_root)
- Metals (D5) are *used_for* framing → 05100_1
- Structural is a *kind_of* "metal framing" (05100_1) → 05100
- Cold formed is a *kind_of* "metal framing" (05100_1) → 05400
- Studes are *made_of* Metals (D5) → (05410_1)
- "Load bearing metal studs" are *kind_of* Metal studs (05410_1) → 05410
- 05410 is *used_for* 05400 → (05400_05410)

Step 3: normalization

- redundancy elimination
- conflict detection
- implication detection

Step 4: generalization

- Synthesize subjects/objects into higher-level concepts connected by the same set of relations



{metals, wood, plastics ..} are *instance_of* Material
{stud, joist ..}are *instance_of* Item
{framing, ..}are *instance_of* Function
{cold formed, structural ..} are *instance_of* Process

Linguistic processing

- inflection, derivation, compounds, and synonyms

ONTOLOGY-BASED SEMANTIC EXTRACTION

- Linguistic Processing such as chunk parsing, and grammatical function recognition [4]
- Matching between relation statements and text

B2010 Exterior Wall:
1. Exterior Wall ***Framing****:* ***Cold-formed****,*
light gage ***steel studs****, C-shape,*
galvanized finish, 6" metal thickness

05100_1 → 05400 → **05410 Load bearing metal studs**

MEASUREMENT OF AFFINITY

- Number of relation statements matched
- Number of keywords matched
- Quality of matches
  1. positions in taxonomy
  2. information content: Inverse document frequency (IDF) [3]
  3. counts in taxonomy

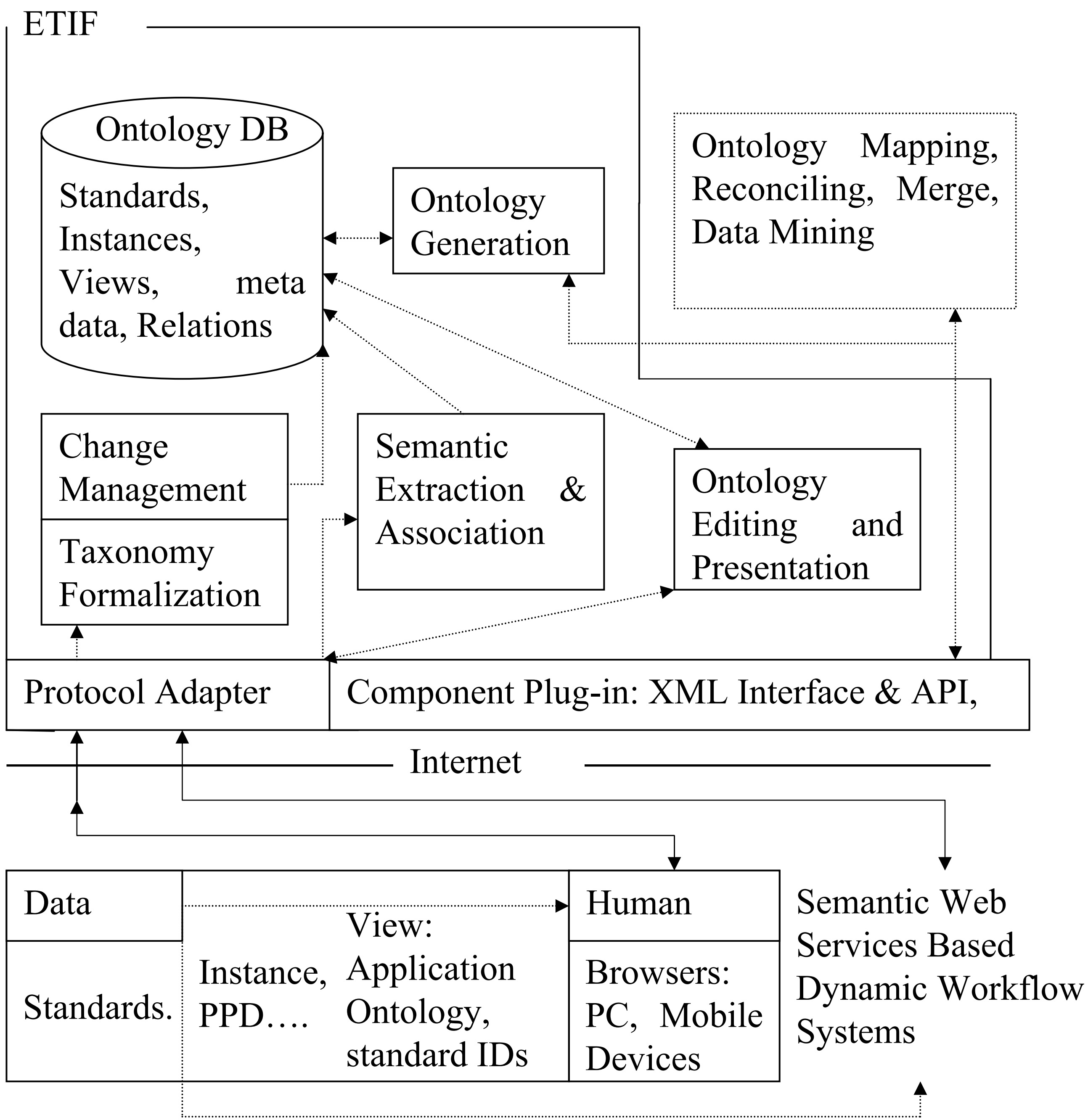


Figure 2. Extensible Taxonomy-based Integration Framework (ETIF)

Conclusion

- Illustration of effective use of taxonomy for improving interoperability in a workflow system with building construction as the target example.
- Illustration of a systematic approach to semantic association of complex complementary taxonomies through knowledge discovery from associated specification documents.

Future Works

- Refinement of the affinity measure
- Integration of the algorithms with dynamic workflow systems through semantic web services.

Acknowledgements

- This work is partially supported by an NSF research grant ITR-0404113

References

- [1] Construction Specifications Institute. MasterFormat 95™ : Alexandria, VA: The Construction Specifications Institute, 1995 edition.
- [2] Charette, R. P. and Marshall, H. E.: UNIFORMAT II Elemental Classification for Building Specifications, Cost Estimating, and Cost Analysis, NISTIR 6389, Gaithersburg, MD: National Institute of Standards and Technology, October, 1999
- [3] Church, K. W. and Gale, W. A. : Inverse document frequency (IDF): A measure of deviations from Poisson. In Yarowsky, D. and Church, K., editors, Proceedings of the Third Workshop on Very Large Corpora, pages 121--130. Association for Computational Linguistics. 1995.
- [4] Maedche, A., Neumann, G., Staab, S.: Bootstrapping an Ontology-Based Information Extraction System, Intelligent Exploration of the Web, Springer 2002.

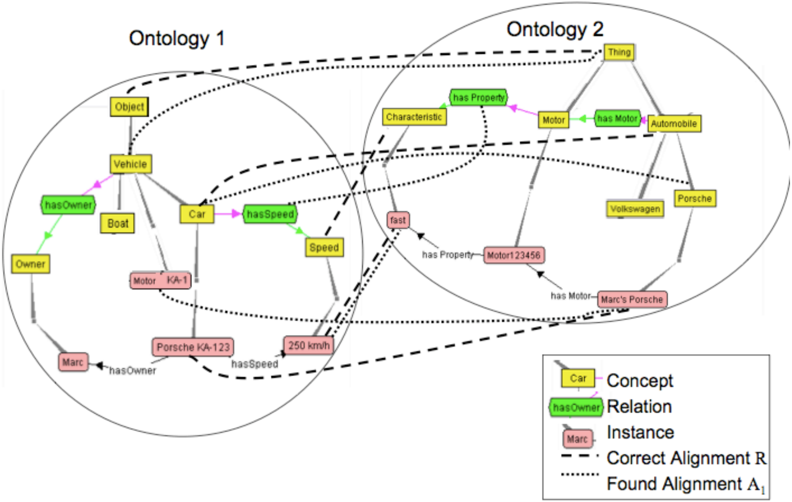
Relaxed precision and recall for ontology matching

Marc Ehrig[1] Jérôme Euzenat[2]

[1] AIFB
University of Karlsruhe
Karlsruhe, Germany
ehrig@aifb.uni-karlsruhe.de

[2] INRIA RHÔNE-ALPES
Monbonnot, France
Jerome.Euzenat@inrialpes.fr

Alignments


Ontology 1
Ontology 2
Object
Vehicle
hasOwner
Boat
Car
hasSpeed
Speed
Owner
Motor KA-1
Porsche KA-123
250 km/h
Marc
hasOwner
hasSpeed
Thing
has Property
Characteristic
Motor
has Motor
Automobile
Volkswagen
Porsche
fast
has Property
Motor123456
has Motor
Marc's Porsche
Car
Concept
hasOwner
Relation
Marc
Instance
Correct Alignment R
Found Alignment A1

Alignments

Definition (Alignment, correspondence)

Given two ontologies O and O', an alignment between O and O' is a set of correspondences (i.e., 4-uples): $\langle e, e', r, n \rangle$ with

- $e \in O$ and $e' \in O'$ being the two matched entities,
- r being a relationship holding between e and e', and
- n expressing the level of confidence [0..1] in this correspondence.

Precision and recall

Definition (Precision, Recall)

Given a reference alignment R, the precision of some alignment A is given by

$$P(A, R) = \frac{|R \cap A|}{|A|}$$

and recall is given by

$$R(A, R) = \frac{|R \cap A|}{|R|}.$$

Problem with precision and recall

ω	(R, R)	(R, A_1)	(R, A_2)	(R, A_3)
precision	1.0	0.2	0.25	0.2
recall	1.0	0.2	0.2	0.2

It does not make a difference between a nearly correct alignment (A_1 or A_2) and a bad one (A_3).

Outline

Solution

- Measuring the "nearly".
- through generalizing precision and recall.

Precision and recall - Generalized

Definition (Generalized precision and recall)

Given a reference alignment R and an overlap function ω between alignments, the precision of an alignment A is given by

$$P_\omega(A, R) = \frac{\omega(A, R)}{|A|}$$

and recall is given by

$$R_\omega(A, R) = \frac{\omega(A, R)}{|R|}.$$

Summary

The main constraint faced by the proximity is the following:

$$|A \cap R| \leq \omega(A, R) \leq min(|A|, |R|)$$

This is indeed a true generalization because, $|A \cap R|$ satisfies all these properties.

Overlap proximity

Definition (Overlap proximity)

The overlap proximity ω between two sets A and R is defined by:

$$\omega(A,R) = \sum_{\langle a,r\rangle \in M(A,R)} \sigma(a,r)$$

in which $M(A,R)$ is a matching between the elements of A and R and $\sigma(a,r)$ a proximity function between two elements.

Choice: the structure of the function.

Matching correspondences

- A matching between alignments is a set of correspondence pairs, i.e., $M(A, R) \subseteq A \times R$.
- We restrict to matchings in which an entity from the ontology does not appear twice. $|M(A, R)| \leq min(|A|, |R|)$.

In precision and recall any correspondence is identified only with itself.
The natural choice is to select the best match because this guarantees that this function generalizes precision and recall.

Best match

Definition (Best match)

The best match $M(A,R)$ between two sets of correspondences A and R, is the subset of $A \times R$ in which each element of A (resp. R) belongs to only one pair, which maximizes the overall proximity:

$$M(A,R) \in Max_{\omega(A,R)}\{M \subseteq A \times R\}$$

Choice: 1-1 match

Correspondence proximity

σ measures the proximity between two matched correspondences:

$$\sigma : M(A, R) \longrightarrow [0\ 1]$$

$$\sigma(\langle e_a, e'_a, n_a, r_a\rangle, \langle e_r, e'_r, n_r, r_r\rangle) = Aggr(\quad \sigma_{pair}(\langle e_a, e_r\rangle, \langle e'_a, e'_r\rangle), \sigma_{rel}(r_a, r_r), \sigma_{conf}(n_a, n_r))$$

We will only consider normalized proximities, i.e., measures whose value ranges within the unit interval [0 1], because this is a convenient way to guarantee that

$$omega(A, R) \leq min(|A|, |R|)$$

Constraints on the aggregation function

The constraints on the aggregation function ($Aggr$) are:

normalization preservation if $\forall i, 0 \leq c_i \leq 1$ then $0 \leq Aggr_i c_i \leq 1$;

maximality if $\forall i, c_i = 1$ then $Aggr_i c_i = 1$;

local monotonicity if $\forall i \neq j, c_i = c'_i = c''_j$ and $c_j \leq c'_j \leq c''_j$ then $Aggr_i c_i \leq Aggr_i c'_i \leq Aggr_i c''_i$.

Correspondence proximity

Definition (Correspondence proximity)

Given two correspondences $\langle e_a, e'_a, r_a, n_a \rangle$ and $\langle e_r, e'_r, r_r, n_r \rangle$, their proximity is:

$$\sigma(\langle e_a, e'_a, r_a, n_a \rangle, \langle e_r, e'_r, r_r, n_r \rangle) =$$

$$\sigma_{pair}(\langle e_a, e_r \rangle, \langle e'_a, e'_r \rangle) \times \sigma_{rel}(r_a, r_r) \times \sigma_{conf}(n_a, n_r)$$

Choice: multiplication as aggregation

Precision/Recall complies to our constraints

Definition (Equality proximity)

The equality proximity is charaterized by:

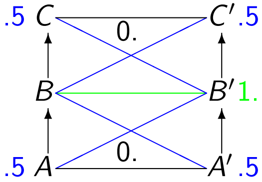
$$\sigma_{pair}(\langle e_a, e'_a\rangle, \langle e_r, e'_r\rangle) = \begin{cases} 1 & \text{if } \langle e_a, e'_a\rangle = \langle e_r, e'_r\rangle \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_{rel}(r_a, r_r) = \begin{cases} 1 & \text{if } r_a = r_r \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_{conf}(n_a, n_r) = \begin{cases} 1 & \text{if } n_a = n_r \\ 0 & \text{otherwise} \end{cases}$$

Symmetric measure

- If the found object is a direct subclass, superclass, subproperty, superproperty, of the expected one, then the proximity will be .5, 0 otherwise.
- If the found relation is $\leq$ instead of $=$, then the proximity is also .5.



This is a fully symmetric measure (i.e., $\omega(A, R) = \omega(R, A)$).

Symmetric measure

Definition (Symmetric proximity)

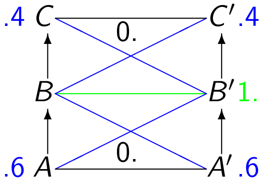
The symmetric proximity is characterized by:

$$\sigma_{pair}(\langle e_a, e'_a\rangle, \langle e_r, e'_r\rangle) \text{ as defined in Table 1}$$
$$\sigma_{rel}(r_a, r_r) \text{ as defined in Table 2}$$
$$\sigma_{conf}(n_a, n_r) = 1 - |n_a - n_r|.$$

Correction effort measure

Measures the effort required by a user to correct an incorrect alignment.

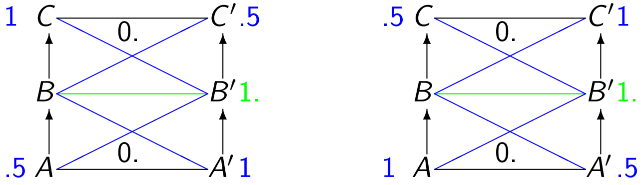
- edit distance-like in which we count the number of operations required for correcting the error.
- very related to the kind of alignment editor available.



The measure is not symmetric because it is easier to change some class for its superclass (very often only one) than for one of its subclasses.

Oriented measure

Different errors will have different impact on the correctness and completeness of answers to an instance retrieving system.
For instance, if instead of an expected class, the alignment find a superclass (in the target ontology), the result will not affect recall (more answers will be returned) but will affect precision.
We use two different ω oriented towards measuring the impact on precision or recall.



This measure is not symmetric.

Evaluation on a simple example

ω	(R,R)		(R,A_1)		(R,A_2)		(R,A_3)	
	P	R	P	R	P	R	P	R
standard	1.0	1.0	0.2	0.2	0.25	0.2	0.2	0.2
symmetric	1.0	1.0	0.4	0.4	0.375	0.3	0.2	0.2
edit	1.0	1.0	0.44	0.44	0.35	0.28	0.2	0.2
oriented	1.0	1.0	0.5	0.5	0.375	0.4	0.2	0.2

Conclusion

We introduced a framework for generalizing precision and recall.
We defined 3(+2) measures implementing this framework.

- they keep precision and recall untouched for the best alignment;
- they help discriminating between irrelevant alignments and not far from target ones;
- specialized measures are able to emphasize some characteristics of alignments: ease of modification, correctness or completeness.

Limitations

- syntactic flavour: semantically equivalent alignments will not be considered the same.
- There has been quite some choices made (see Choice mentions).
- Some general principles to choose weights are required.

Questions?

ehrig@aifb.uni-karlsruhe.de
Jerome.Euzenat@inrialpes.fr

Searching Web Resources Using Ontology Mappings

Dragan Gašević and Marek Hatala
Laboratory for Ontological Research
School of Interactive Arts and Technology
Simon Fraser University Surrey, Canada
http://lore.iat.sfu.ca
Email: {dgasevic, mhatala}@sfu.ca

Outline

- Introduction
- Metadata, Ontologies, and Web resources
- Representation of ontologies and mappings
- Ontology mapping based search algorithm
- Evaluation
- Conclusions

Introduction

- Collections of web resources:
  - digital libraries, community-based object repositories, dispersed web resources in many individual institutions
- Resources
  - typically not interconnected into the web
- Interoperability
  - subject categories, taxonomies, ideally richer ontologies
  - diversity among institutions
- Ontology mapping as a solution

Web resource metadata and domain ontologies

- Description of web resources
  - Metadata schema: Dublin Core (DC)
  - Predefined set of relevant fields
  - DC is defined as an RDF Schema
  - Resource are annotated with the schema instances
    - instance (individual) level

Web resource metadata and domain ontologies

- Semantic annotations
  - From ontologies
    - RDFS, OWL,…
    - Classifications as ontology schemas
      - Dewey Decimal System, ACM CCS, directories, …
  - Complement metadata
    - Metadata are ontology schema instances
    - Classification are ontology schemas
    - Conceptual mismatch
      - Possible in OWL
      - Problems with implementation

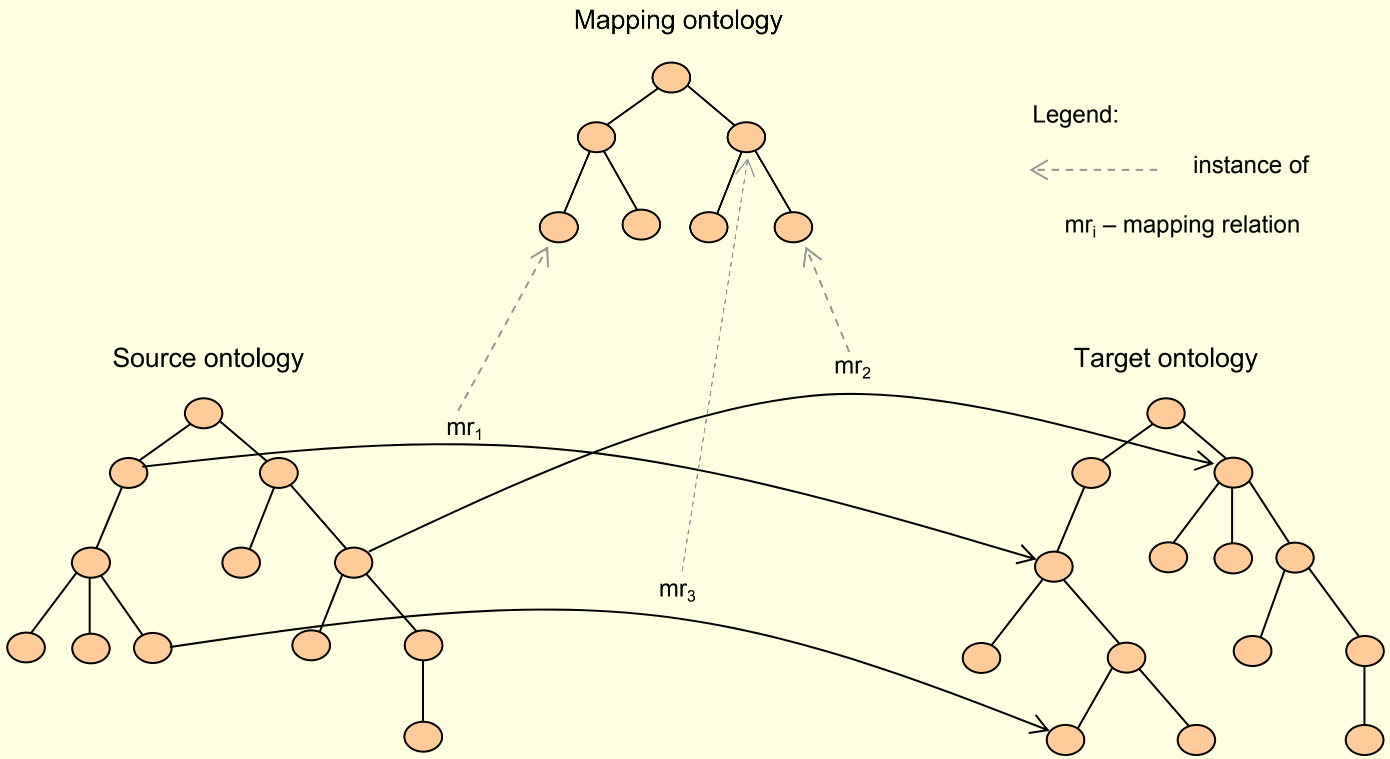
Web resource metadata and domain ontologies

- Multiple domain ontologies or classifications
  - Different systems have different needs
    - Digital libraries – library classifications
    - Ontologies or taxonomies for domain specific application
    - …
- How to search multiple web resource collections based on multiple classifications/taxonomies//domain ontologies?

Multiple domain ontologies

- Ontology mappings
  - Define how concepts from different ontologies relate each other
- Mapping ontology
  - Reusable problem-solving components [Crubézy et al., 2003]
    - Mapping ontology between domain and method ontologies
  - MApping FRAmework (MAFRA) [Maedche et al, 2002]
    - Semantic bridge ontology

Mapping ontology approach



Multiple domain ontologies

- Ontology mappings
  - Define how concepts from different ontologies relate each other
- Mapping ontology
  - Reusable problem-solving components [Crubézy et al., 2003]
    - Mapping ontology between domain and method ontologies
  - MApping FRAmework (MAFRA) [Maedche et al, 2002]
    - Semantic bridge ontology
- Summary
    - There is no widely accepted solution
    - Different mapping types

Our approach

- Representation of ontologies and ontology mappings
  - Simple Knowledge Organization System (SKOS)
    - A recent RDF(S)/OWL-based W3C effort
  - Three vocabularies:
    - *SKOS Core*
    - *SKOS Extensions*
    - *SKOS Mapping*

SKOS Core

SKOS Core	Class/Property	Description
ConceptScheme	Class	A set of concepts, optionally including statements about semantic relationships between those concepts.
Concept	Class	A resource is a conceptual resource.
inScheme	Property	A concept is a part of a particular concept scheme
hasTopConcept	Property	A link between the concept scheme and the concepts that are the top-level concepts in the generalization hierarchy.
prefLabel and *altLabel*	Property	Preferred and alternative lexical labels of a resource.
broader	Property	A concept is broader in meaning (i.e. more general) than another.
narrower	Property	A concept is narrower in meaning (i.e. more specific) than another Inverse to the *broader* property. Transitive property.

- SKOS Extensions
  - *narrowerGeneric* and *broaderGeneric* are subproperties of *narrower* and *broader*, respectively
  - Equivalent to *rdfs:subPropertyOf*

SKOS Mappings

- Rich set of mapping properties

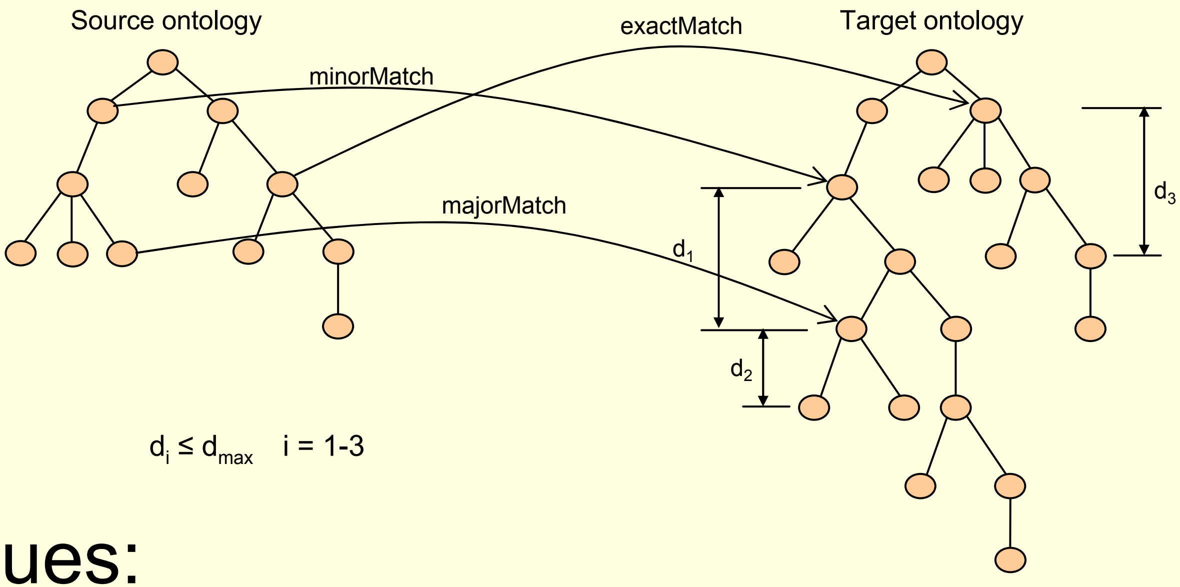
SKOS Mapping Property	Description
mappingRelation	The super-property of all properties expressing information about how to create mappings between concepts from different conceptual schemes.
broadMatch	The set of resources properly indexed against the first concept is a subset of the set of resources properly indexed against concept the second concept.
narrowMatch	The set of resources properly indexed against the first concept is a superset of the set of resources properly indexed against concept the second concept.
exactMatch	The set of resources properly indexed against the first concept is identical to the set of resources properly indexed against the second.
majorMatch	The first concept shares more than 50% of its members with the set of resources properly indexed against the second concept.
minorMatch	The set of resources properly indexed against the first concept shares less than 50% but greater than 0 of its members with the set of resources properly indexed against the concept.

Ontology mapping based search algorithm

- Multiple ontologies
- Mapping ontology defines relations between ontologies
- Observed case:
  - Two ontologies (source and target), but the algorithm is not limited to just two solution
  - Input are concepts of the source ontology
  - Results are concepts of the target ontology
  - Different mapping relations have different influence on ranking

Ontology mapping based search algorithm

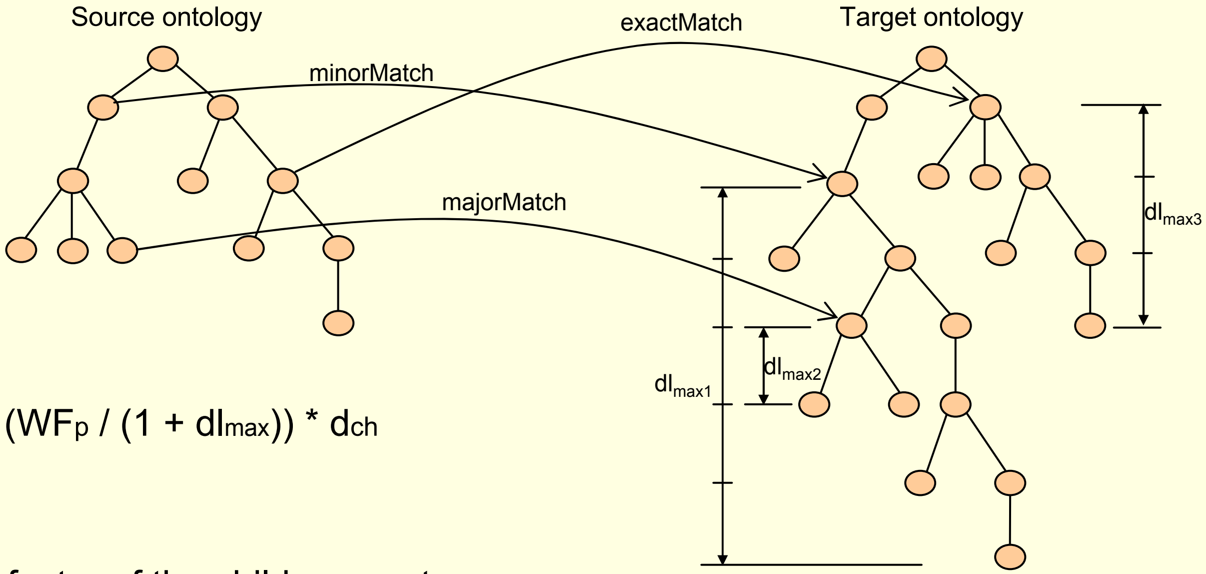
- Initial version
  - Depth-limited search (d_{max})



- Issues:
  - the resulting concept list is completely discrete
  - some relevant child concepts can be taken out of consideration due to depth limit

Ontology mapping based search algorithm

- Improved version



- $WF_{ch} = WF_p - (WF_p / (1 + dl_{max})) * d_{ch}$

- WF_{ch} – weight factor of the child concept;
- WF_p – weight factor of the matched (parent) concept;
- dl_{max} – maximal depth level of the matched (parent) concept;
- d_{ch} – distance of the child concept from the matched (parent) concept

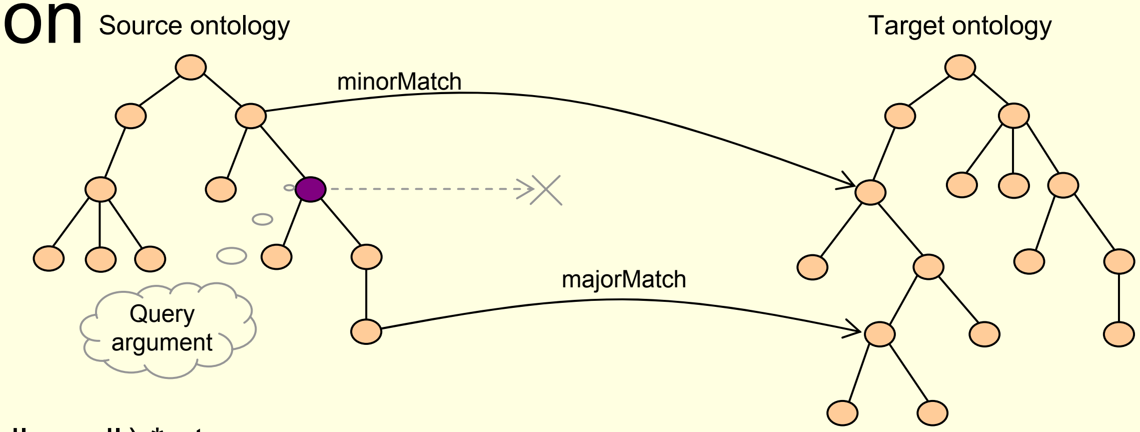
Ontology mapping based search algorithm

- Improved version

```
function search-concept (input-concept, WFEM)
     cluster-names := {"exactMatch", "broadMatch", "exactMatchChildren",
     "broadMatchChildren" , "narrowMatch", "narrowMatchChildren" "majorMatch",
     "majorMatchChildren", "minorMatch", "minorMatchChildren"};

     clusters := create-hash-map();
     Result   := {};

     for-each name in cluster-names
          matched-concepts := get-matched-concepts(name, input-concept);
          clusters[name] := matched-concepts;
     end-for-each

     for-each name in cluster-names
          for-each concept in clusters[name]
                    put-in-sorted-list(result, concept, calculate-WF(concept, name));
          end-for-each
     end-for-each

     return result;
end-function
```

Ontology mapping based search algorithm

- Final version



- $WF_i = WF_{EM} - abs(dl_{sc} - dl_i) * step$

- WF_{EM} – weight factor of the exact match relation predefined for the case when there is a mapping relation between the query argument and the target ontology;
- dl_{sc} – depth level of the query argument;
- dl_i – depth level of a parent/child concept of the query argument that has a mapping relation with the target ontology;
- step – predefined value that specifies the impact of the distance between the query argument and its child/parent concept *i*.

Ontology mapping based search algorithm

- Final version

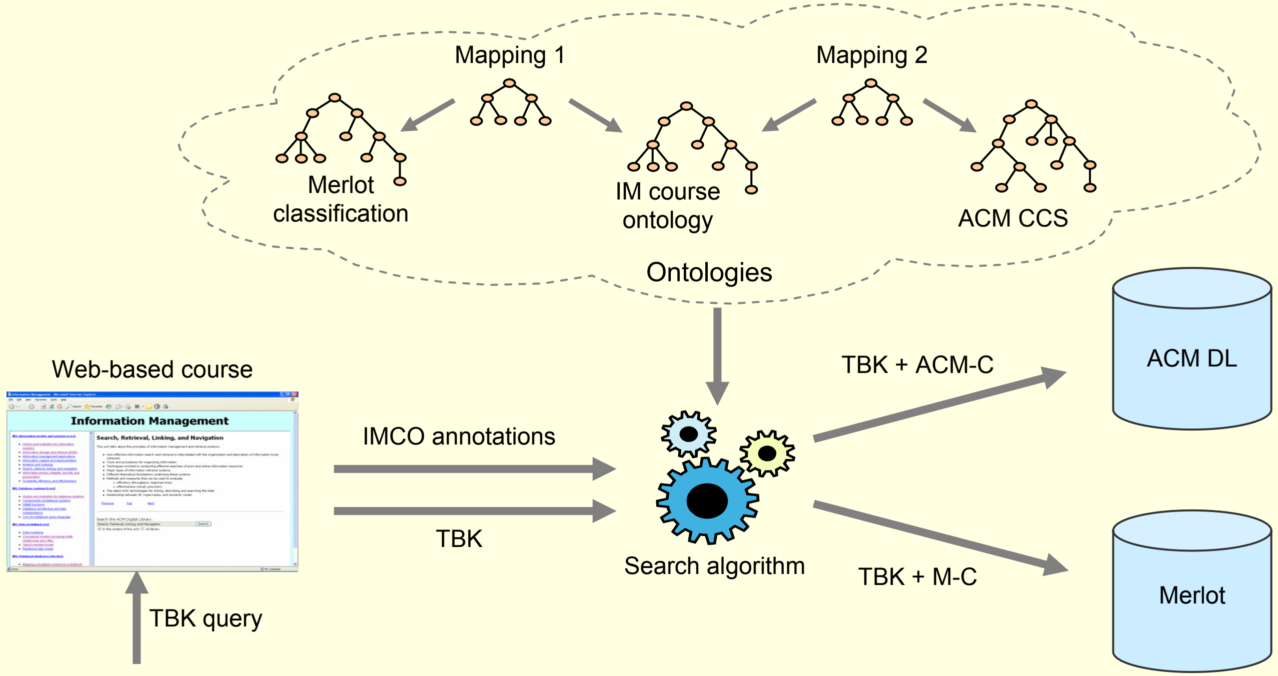
```
function search-concept-no-direct-match (input-concept, WFEM)
     result := search concept(input-concept, WFEM);
     if result == {} then
          children := get-subconcepts-with-mapping(input-concept);
          parents := get-superconcepts-with-mapping(input-concept);

          for-each c in children
                    WF := calculate-WF(c, input-concept);
                    put-in-ordered-list(result, search-concept(input-concept, WF));
          end-for-each

          for-each c in parents
                    WF := calculate-WF(c, input-concept);
                    put-in-ordered-list(result, search-concept(input-concept, WF));
          end-for-each
     end-if

     return result;
end-function
```

Implementation

- Algorithm implementation
  - Jess and OWLJessKB
  - Component that can be used in different applications

Evaluation

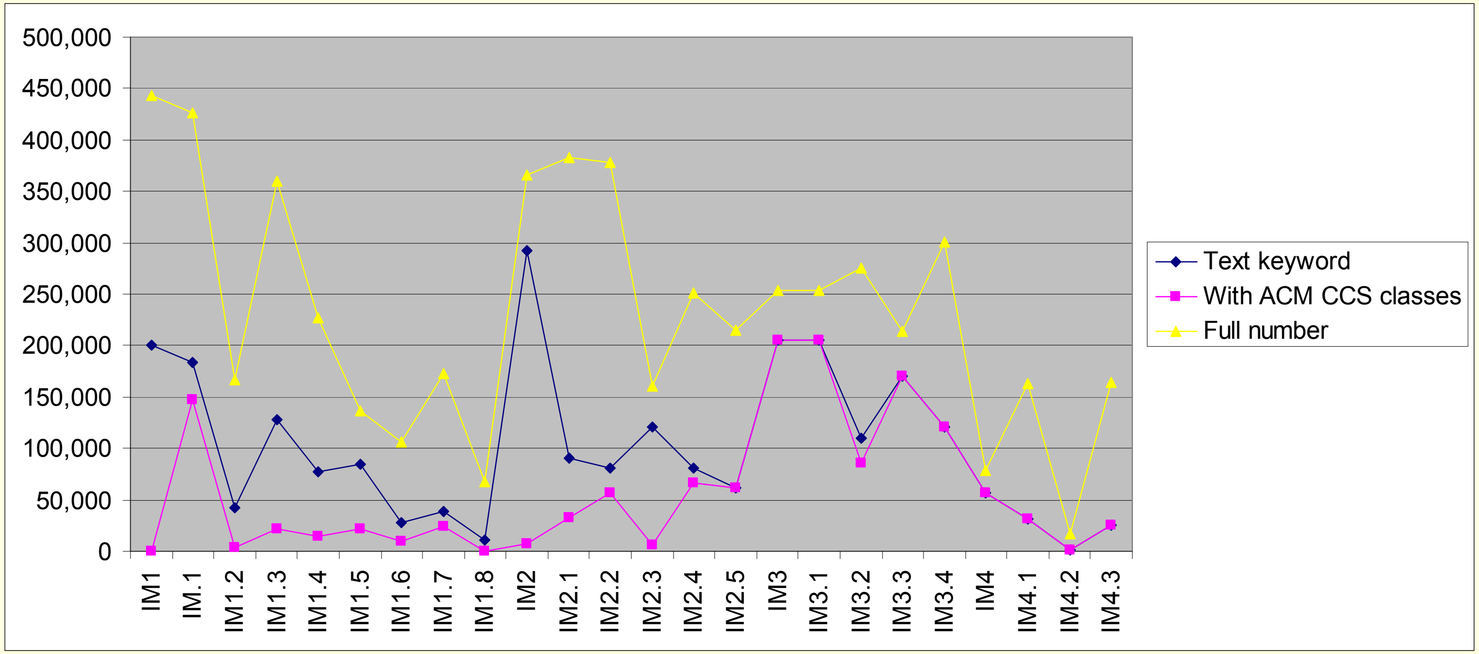
- Evaluation environment



Legend:
- IMCO – IM course ontology
- TBK – text-based keywords
- M-C – classifiers from the Merlot classification
- ACM-C – classifiers form the ACM CCS

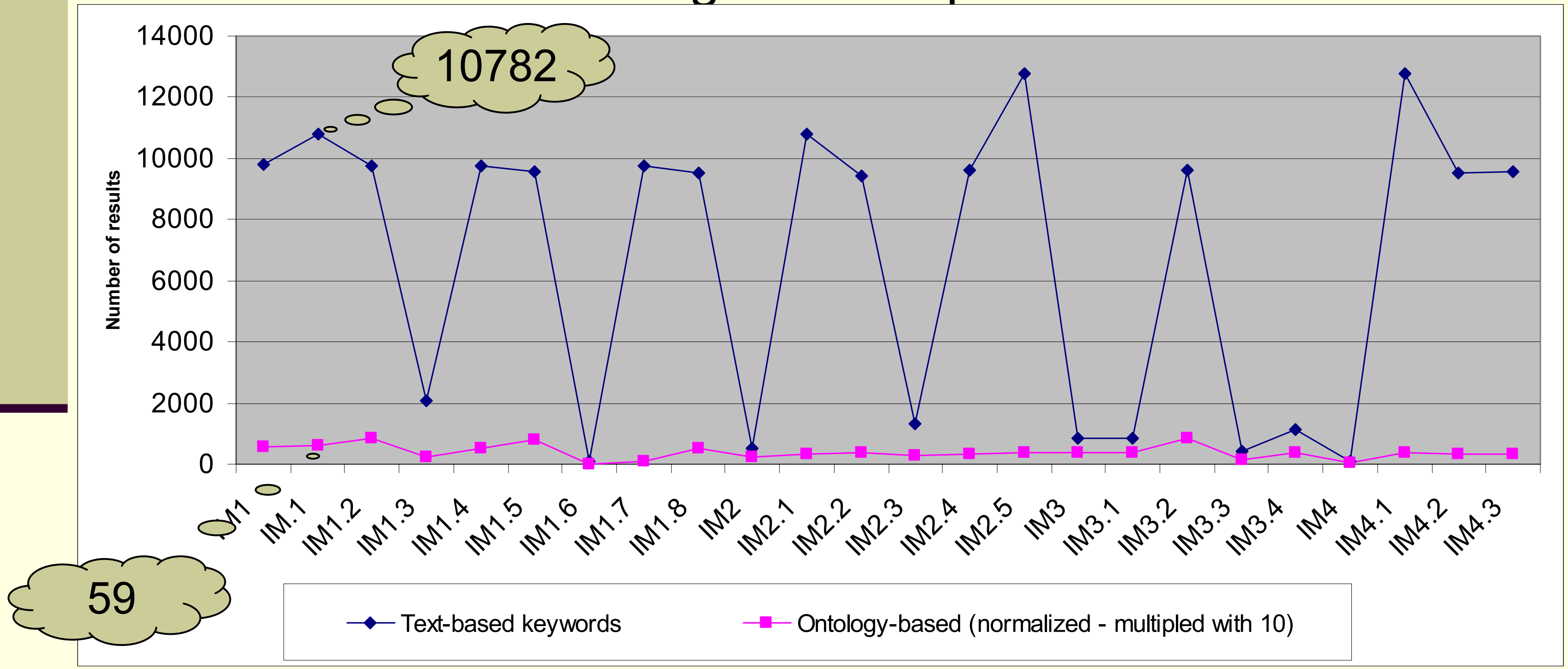
Evaluation

- ACM Digital Library
  - Results opposite to our expectations
    - OR operator did not give expected results
      - Verity search engine
      - Threshold
    - Big number of classifiers decreases the result set



Evaluation

- Merlot – Learning object repository
  - Merlot classification system – general purpose one
  - Results according to our expectations



Evaluation

- Merlot – Learning object repository

Concept	IM1	IM.1	IM1.2	IM1.3	IM1.4	IM1.5	IM1.6	IM1.7	IM1.8	IM2	IM2.1	…
Keyword-based search	9814	10782	9760	2094	9769	9578	114	9760	9542	540	10797	…
Ontology-based search	55	59	85	22	53	80	1	9	52	25	35	…
Percent	0.56	0.55	0.87	1.05	0.54	0.84	0.88	0.09	0.54	4.63	0.32	…
Num. of classification tags	1	1	3	2	1	1	1	1	1	2	2	…
Defined match or not	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N	…

Concept	…	IM2.2	IM2.3	IM2.4	IM2.5	IM3.2	IM3.3	IM3.4	IM4	IM4.1	IM4.2	IM4.3
Keyword-based	…	9449	1321	9638	12782	9614	418	1140	72	12788	9544	9563
Ontology-based	…	36	26	35	38	85	14	40	6	38	31	31
Percent	…	0.38	1.97	0.36	0.30	0.88	3.35	3.51	8.33	0.30	0.32	0.32
Num. of classification tags	…	2	2	2	2	3	3	3	2	2	2	2
Defined mapping or not	…	N	N	N	N	N	N	N	Y	N	N	N

Conclusions and future work

- The ontology mappings algorithm
  to get semantically relevant search results
- Initial evaluation results are promising
- In the future:
  - Evaluation with other similar approaches
  - eduSource Communication Layer (ECL)
    federated search engine
  - Examining on the OWL language
  - Improving ranking algorithm –
    different influence of different properties
  - Automatic mapping discovery using semantic signatures

Searching Web Resources Using Ontology Mappings

Dragan Gašević and Marek Hatala
Laboratory for Ontological Research
School of Interactive Arts and Technology
Simon Fraser University Surrey, Canada
http://lore.iat.sfu.ca
Email: {dgasevic, mhatala}@sfu.ca

GMO: A Graph Matching for Ontologies

Wei Hu, Ningsheng Jian, Yuzhong Qu & Yanbing Wang

Dept. Computer Science & Engineering

Southeast University, China

Outline

- Introduction
- RDF Bipartite Graph
- Structural Similarity for Ontology
- Implementation
- Experimental Results
- Discussion

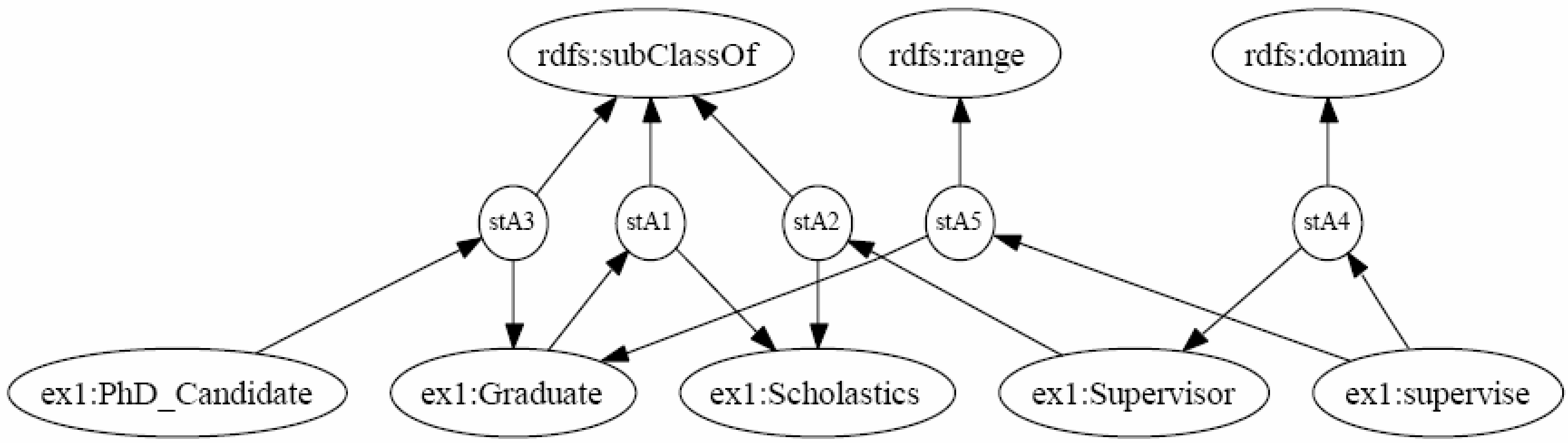
XOBJECTS
XOBJECTS.SEU.EDU.CN

Introduction

- **Two Distinguished Features**
  - **It uses directed bipartite graphs (statement vs. entity) to represent ontologies instead of using labeled graphs or RDF graphs.**
  - **A new measure of structural similarity for web ontologies. This measure will play an important role in ontology matching, especially when lexical similarity could not be gained.**
- **One of the main components in Falcon-AO**

Directed RDF Bipartite Graph

- RDF Bipartite Graph Model
- Directed Bipartite Graph



Matrix Representation

- The Adjacent Matrix of Ontology

$$A = \begin{pmatrix} 0 & 0 & A_{ES} \\ 0 & 0 & A_S \\ A_E & A_{OP} & 0 \end{pmatrix}$$

The Idea of Our Measure

- Similarity of two entities from two ontologies comes from the accumulation of similarities of involved statements (triples) taking the two entities as the same role (subject, predicate, object) in the triples.
- Similarity of two statements comes from the accumulation of similarities of involved entities (including external entities) of the same role in the two statements being compared.

Structural Similarity for Ontology

- The updating equations

$$
\begin{aligned}
O_{k+1} &= B_S S_k A_S^T + B_{OP}^T S_k A_{OP} \\
S_{k+1} &= B_E E_{BA} A_E^T + B_{ES}^T E_{BA} A_{ES} \\
&\quad + B_{OP} O_k A_{OP}^T + B_S^T O_k A_S
\end{aligned}
$$

$$
\begin{aligned}
A &= \begin{pmatrix} 0 & 0 & A_{ES} \\ 0 & 0 & A_S \\ A_E & A_{OP} & 0 \end{pmatrix} \\
B &= \begin{pmatrix} 0 & 0 & B_{ES} \\ 0 & 0 & B_S \\ B_E & B_{OP} & 0 \end{pmatrix} \\
X_k &= \begin{pmatrix} E_{BA} & & \\ & O_k & \\ & & S_k \end{pmatrix}
\end{aligned}
$$

Refinement ❶

- Classify the entities described in a given ontology as properties, classes and instances.

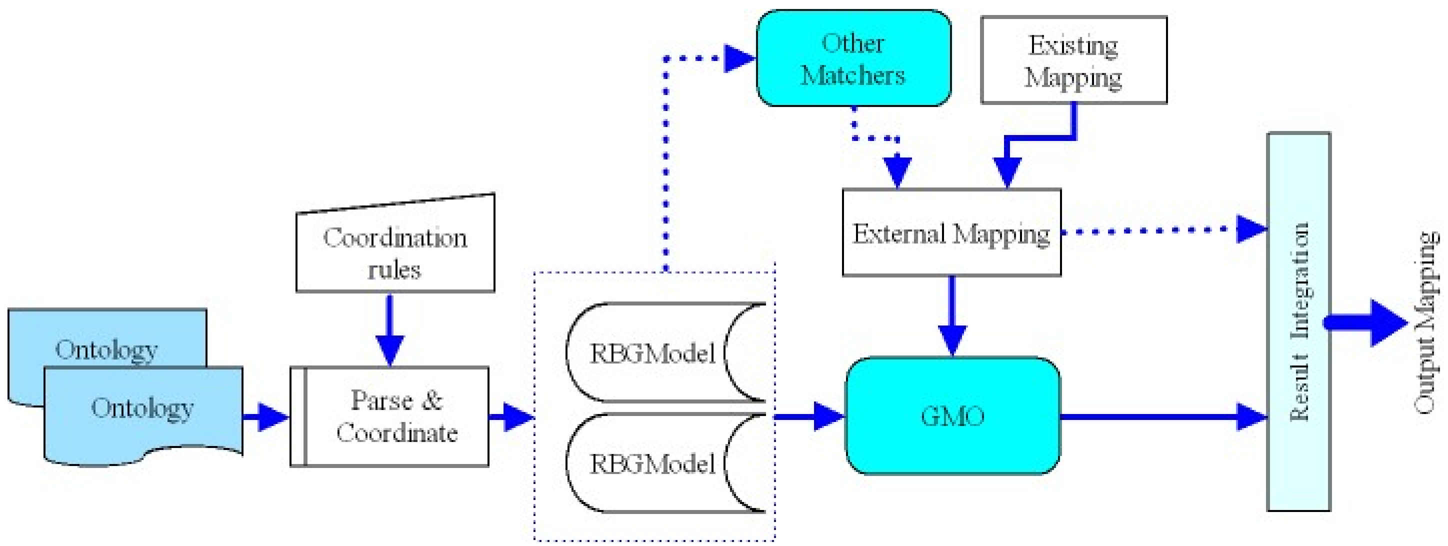
$$O_k = \begin{pmatrix} P_k & & \\ & C_k & \\ & & I_k \end{pmatrix}$$

Refinement ❷

- Two advantages
  - Good computing performance due to the matrix computation with blocks.
  - Avoid the unnecessary computing of similarity between different kinds of entities.

Implementation

❑ Matching Process of GMO



Coordinating Ontologies

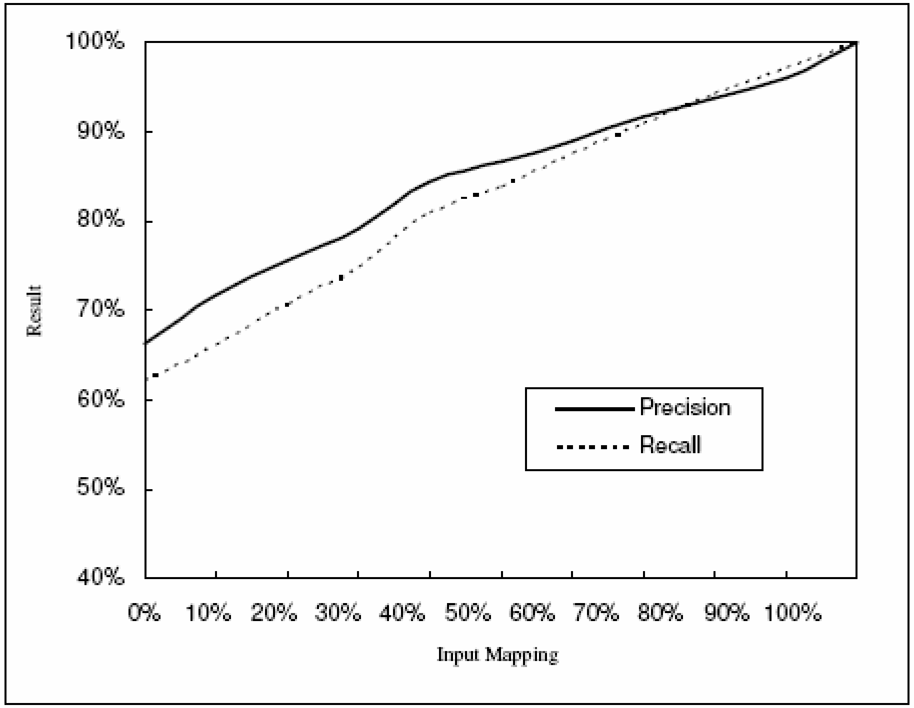
- Discarding (ontology header, etc.)
- Merging (owl:equivalentClass, etc.)
- Inference
- List (rdfs:member)

LMO

- A Linguistic Matching for Ontologies
- Another Matcher in Falcon-AO

Effect of GMO

- We test the effectiveness of GMO on OAEI 2005 benchmark test cases, by taking some percentage of standard matched pairs as input mapping to GMO.



Average Precision & Recall

Performance of Falcon-AO

- The partial experiment results of Falcon-AO

	101-104	201-210	221-266	301-304	Total
Prec.	1.0	0.96	0.86	0.93	0.89
Reca.	1.0	0.95	0.82	0.81	0.85
F-M.	1.0	0.95	0.83	0.86	0.87

Discussion ❶

- Advantages
  - GMO uses directed bipartite graphs to represent web ontologies instead of using labeled graph or RDF graph.
  - Our similarity model emphasizes the structural similarity based on the connection similarity, and does not depend on or mix up with lexical similarity.
  - In addition, GMO can make use of a set of matched pairs found previously by other approaches as external entities.

Discussion ❷

- **Weaknesses**
  - It performs not so well when the ontologies to be matched have a great difference in structure.
  - Sometimes, it is really hard to distinguish the exact mapping only by structural features.
  - It is not easy to select appropriate coordination rules due to the tradeoff between the cost of inference and the quality of mapping.

Main Reference

- V. Blondel, A. Gajardo, M. Heymans, P. Senellart, P. Van Dooren. A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching. *SIAM Review (2004)*
- J. Hayes, C. Gutiérrez. Bipartite Graphs as Intermediate Model for RDF. *ISWC (2004)*
- S. Melnik, H. Garcia-Molina, E. Rahm. Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching. *ICDE (2002)*
- Y. Sure, O. Corcho, J. Euzenat, T. Hughes (eds.). Proceedings of the 3rd International Workshop on Evaluation of Ontology Based Tools (EON 2004). *CEUR-WS Publication (2004)*

Thanks !

Any Comment and Suggestion is Welcome!

CENTRO DE TECNOLOGÍAS DE INTERACCIÓN VISUAL Y COMUNICACIONES

VICOM Tech

VISUAL COMMUNICATION TECHNOLOGIES

VISUAL INTERACTION AND COMMUNICATIONS TECHNOLOGIES

Carlos Lamsfus. K-CAP 2005

Banff, October 2nd 2005

Who are we?

VICOM Tech
VISUAL COMMUNICATION TECHNOLOGIES

- **A non profit Applied Research Technology Centre** in Computer Graphics, Multimedia and Telecommunications.
- Located in **San Sebastian,** Spain, in the San Sebastian Technology Park.
- Founded by the INI-GraphicsNet and EiTB (April 2001)

- R & D Center, integrated in the Basque Technology Network (Saretek) as **Center of Excellence in R&D and Technology Transfer**
- **Member of INI-GraphicsNet**
- About 35 Researchers (Engineers, Computer Scientists, Students, etc)
- VICOMTech is an ISO 9001:2000 certified institute

eitb

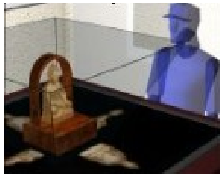
Foundation Stiftung
INI-GraphicsNet

Application Areas

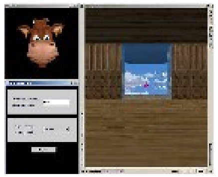
VICOM Tech
VISUAL COMMUNICATION TECHNOLOGIES

- Digital TV and Interactive Services

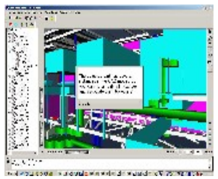
- Medical Applications



- Cultural Heritage & GIS



- Education, Entertainment and e-Learning



- Industrial Applications

Semantic Web Technologies

eitb

Foundation Stiftung
INI-GraphicsNet

Towards Semantic Based Information Exchange and Integration Standards: the art-E-fact ontology as an extension to the CIDOC CRM (ISO/CD 21127) Standard

Carlos Lamsfus, María Teresa Linaza and Tim Smithers

Introduction: The Context

- Information exchange
- Knowledge sharing

⇒ Internet, Ontologies

- Different systems (technical)
- Different culture /methodology/ languages (semantic)

⇒ Interoperability

- Pursue a culture of re-use of already exiting work
- Contribute to standards

⇒ Standards

Presentation Overview

VICOM Tech
VISUAL COMMUNICATION TECHNOLOGIES

- Introduction and Objectives
- Previous work
  - The art-E-fact project and ontology
  - The CIDOC CRM ontology
- art-E-fact vs. CIDOC CRM -> Differences
- Alignment of the art-E-fact and CIDOC CRM ontologies
- Conclusions

eitb

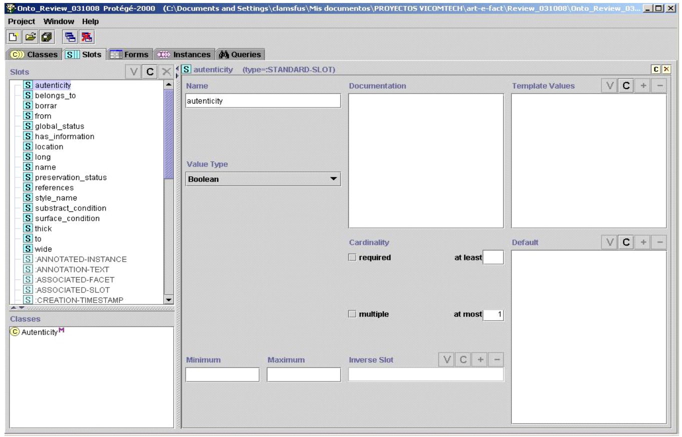
Foundation Stiftung INI-GraphicsNet

Create a generic platform for Interactive Storytelling in Mixed Reality that allows artists to create artistic expressions in an original way within a cultural context between the virtual and the physical reality

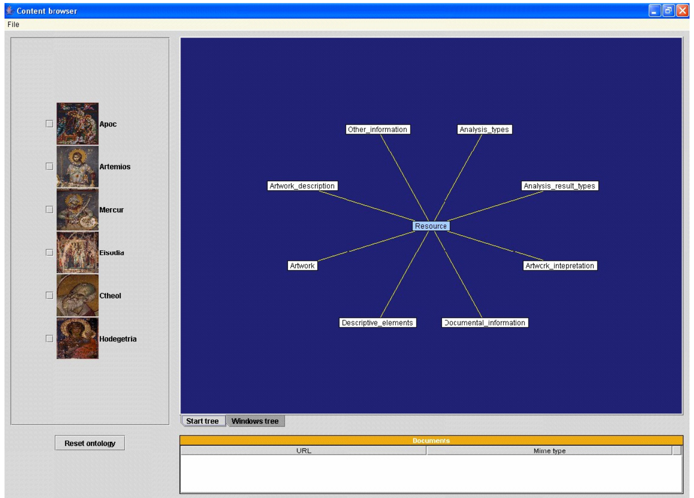
- Develop a generic platform for interactive storytelling
- Facilitate access to a knowledge database of cultural and artistic material
- Develop an Authoring-Tool (from scratch) that allows artists to create interactive stories (content, virtual characters, background and interaction metaphors)
- **Access to the content databases**

Previous work: the art-E-fact ontology

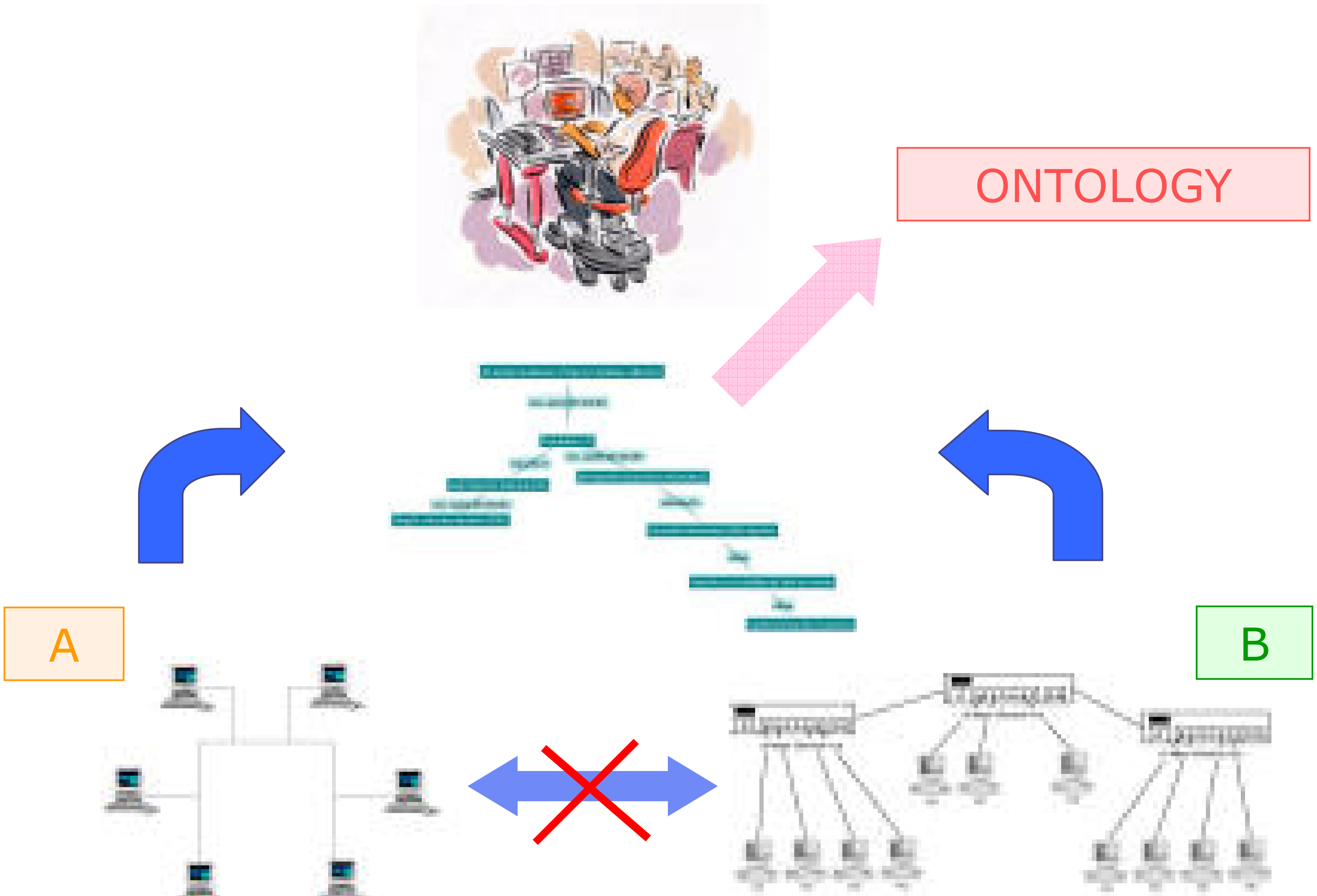
- The art-E-fact ontology:
  - For authors to get a general idea of the content
  - Reflect relations among concepts that are not shown in the database



- The Content Browser:
  - Efficient and effective access and navigation through the concepts
  - To get to know and discover what there is available
  - Access to the content database



Previous work: the CIDOC CRM(I/II)



Previous work: the CIDOC CRM (II/II)

- Serve as common language for domain IT experts and developers

- Support the implementation of automatic data transformation algorithms from local to global structures without loss of meaning

- Exchange and integration of heterogeneous scientific documentation of museum collections:
  - Scientific documentation -> information described by CIDOC CRM as sufficient for academic research
  - Museum collections -> collections, sites, monuments, etc.

Presentation Overview

- Introduction and Objectives
- Previous work
  - The art-E-fact project and ontology
  - The CIDOC CRM ontology
- art-E-fact vs. CIDOC CRM -> Differences
- Alignment of the art-E-fact and CIDOC CRM ontologies
- Conclusions

art-E-fact vs. CIDOC CRM -> differences

Similarities

- Both ontologies reflect a (serious) commitment to the expression of common concepts underlying data structures used by their users

Differences

- The art-E-fact ontology was motivated by the need to describe added-value content for the creation of stories
- The CIDOC CRM ontology focuses on documentation processes among cultural institutions, motivated by the need to share information

art-E-fact vs. CIDOC CRM -> differences

VICOM Tech VISUAL COMMUNICATION TECHNOLOGIES

CIDOC CRM

- SCOPE: all the information required for the scientific documentation of cultural heritage collections -> information exchange
- CIDOC CRM focuses on curated knowledge of museums
- The CIDOC CRM is intended to cover contextual information, e.g. historical, geographical and theoretical background

art-E-fact

- SCOPE: the ontology is not devoted to documentation, but to content description and comprehension -> "semantic index"
- art-E-fact focuses on content generation by artists
- The art-E-fact ontology takes into account different levels of knowledge in order to provide rich content to build interactive stories

eitb

Foundation Stiftung INI-GraphicsNet

Presentation Overview

VICOM Tech VISUAL COMMUNICATION TECHNOLOGIES

- Introduction and Objectives
- Previous work
  - The art-E-fact project and ontology
  - The CIDOC CRM ontology
- art-E-fact vs. CIDOC CRM -> Differences
- **Alignment of the art-E-fact and CIDOC CRM ontologies**
- Conclusions

eitb

Foundation Stiftung INI-GraphicsNet

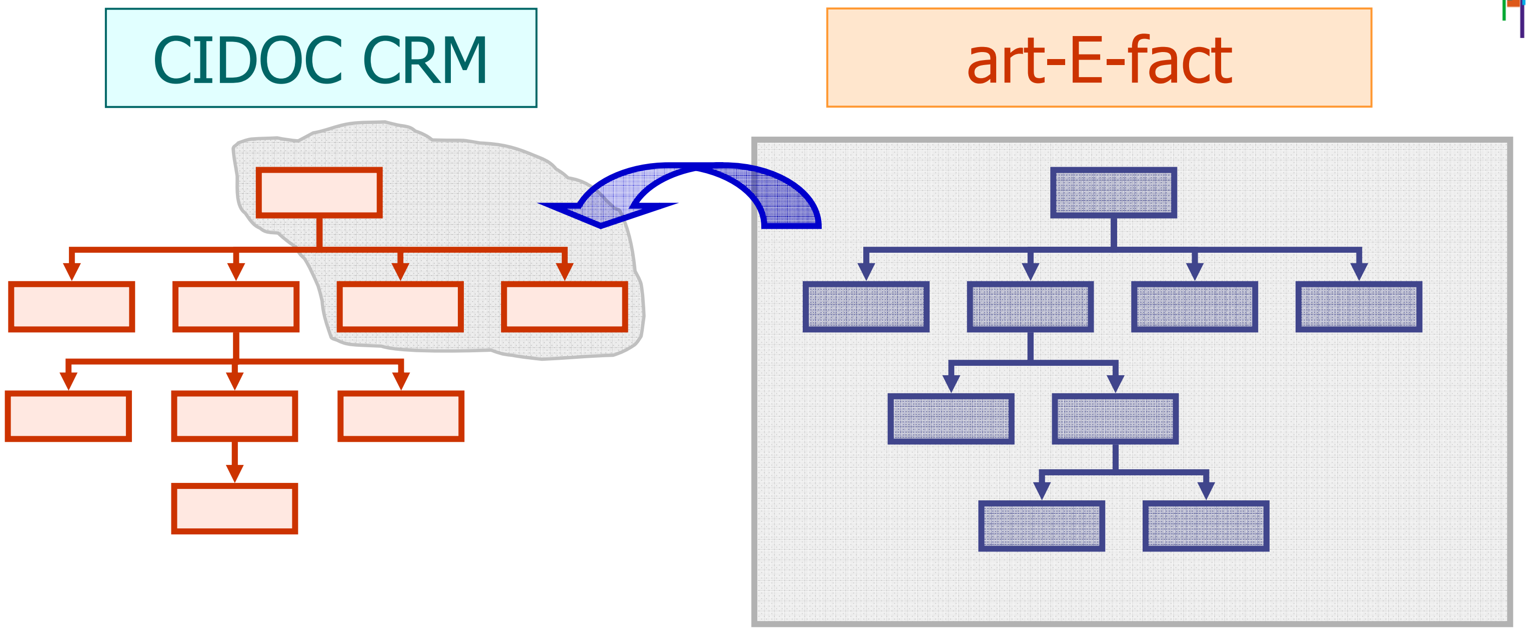
Alignment of the art-E-fact and CIDOC CRM ontologies (I/III)

- Merging vs. Alignment (incorporation) of ontologies

- Questions:
  - Does the art-E-fact ontology need to be a CRM extension?
  - What would we like to do with the extended version?
  - What do we want to support people doing?

- Alignment: semi-automated rule-based process
  - Tool -> to be selected yet
  - Ontology language: OWL DL
  - Alignment language: RWL (http://www.wsmo.org/wsml/wrl/wrl.html#wsml)

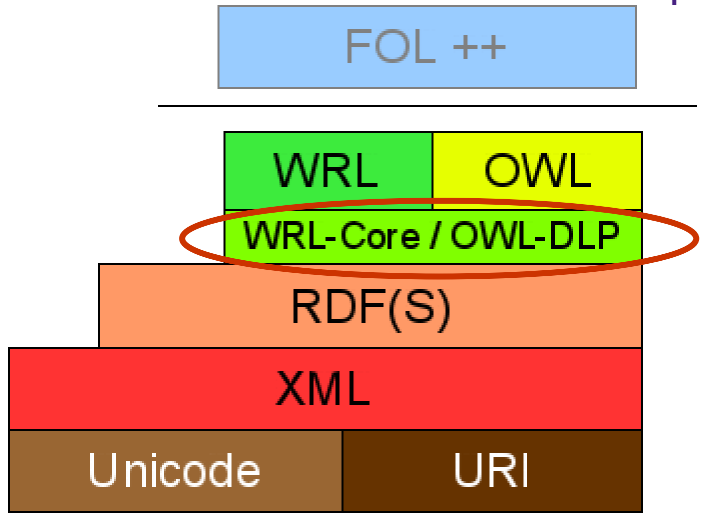
Alignment of the art-E-fact and CIDOC CRM ontologies (II/III)



- Understand how the art-E-fact ontology is related to the CRM (knowledge levels)
- Identify CRM's part we want to map to art-E-fact
- Try to find a CRM subgroup and match it (semantically) as identities

Alignment of the art-E-fact and CIDOC CRM ontologies (III/III)

- WRL (Web-based Rule Language)
  - Derived from the ontology component WSML
  - Rule-based ontology language (OWL: description logic language)



- "Translate" the art-E-fact ontology into OWL DL

- Using WRL identify "common" concepts

Presentation Overview

- Introduction and Objectives
- Previous work
  - The art-E-fact project and ontology
  - The CIDOC CRM ontology
- art-E-fact vs. CIDOC CRM -> Differences
- Alignment of the art-E-fact and CIDOC CRM ontologies
- Conclusions

Conclusions

- Technology tending to standards -> enable information exchange
- The art-E-fact and CIDOC CRM ontologies
  - Definition, comparison, differences -> conclusions
- Research on semantic-based rule languages
- Contribute in general to the standardization of processes as well as to standards
- Concrete example of the application of the mapping process

CENTRO DE TECNOLOGÍAS DE INTERACCIÓN VISUAL Y COMUNICACIONES

VICOM Tech

VISUAL COMMUNICATION TECHNOLOGIES

VISUAL INTERACTION AND COMMUNICATIONS TECHNOLOGIES

Carlos Lamsfus. K-CAP 2005

Banff, October 2nd 2005

An approach to ontology mapping negotiation

Presentation at K-CAP'05 IntOnt WS
Banff, Canada
2005-10-02

Nuno Silva, Paulo Maio, João Rocha

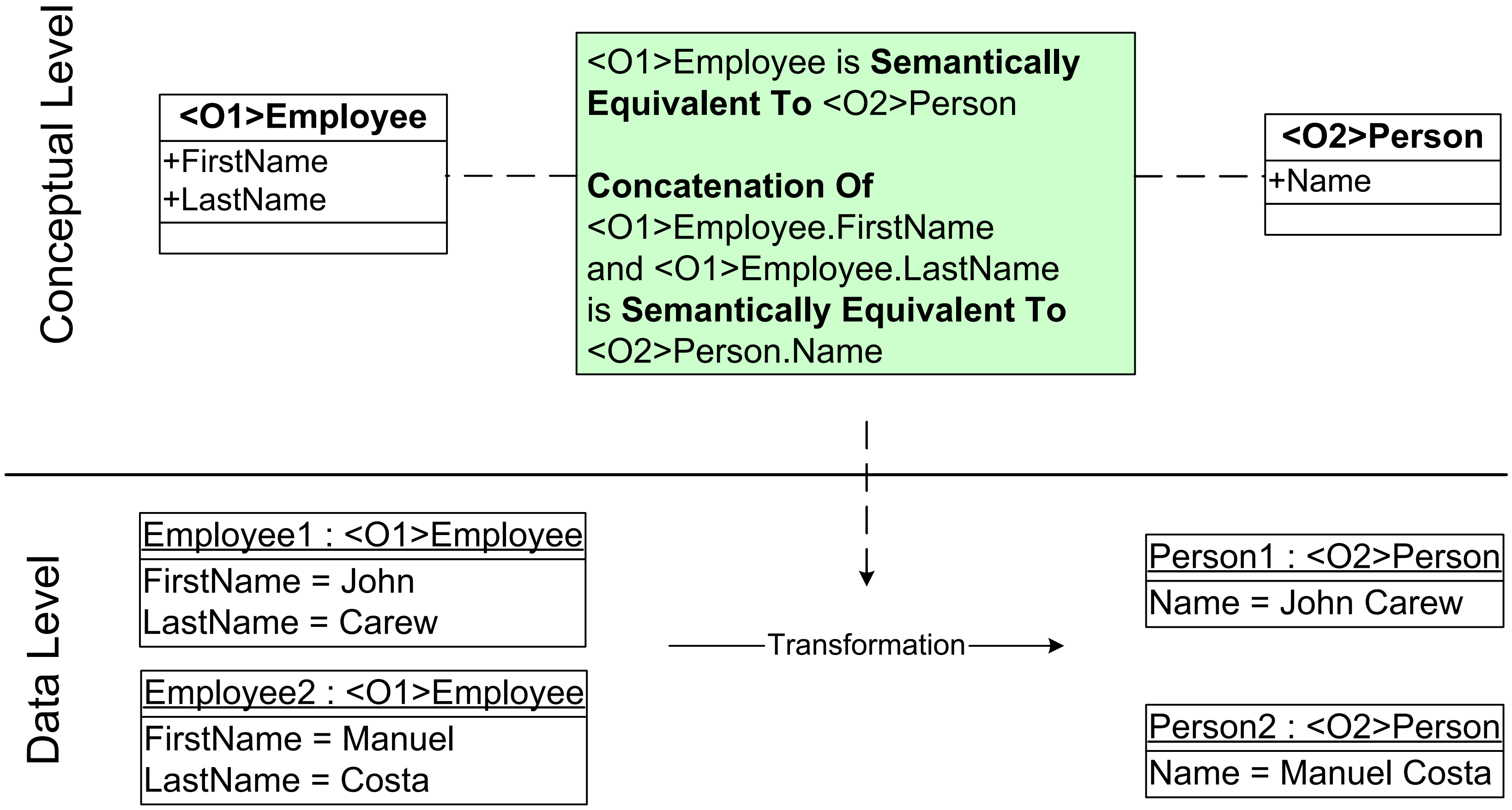
GECAD – Knowledge Engineering and Decision Support Research Group
Polytechnic Institute of Porto - Portugal

http://www.gecad.isep.ipp.pt
Nuno.Silva@dei.isep.ipp.pt

Agenda

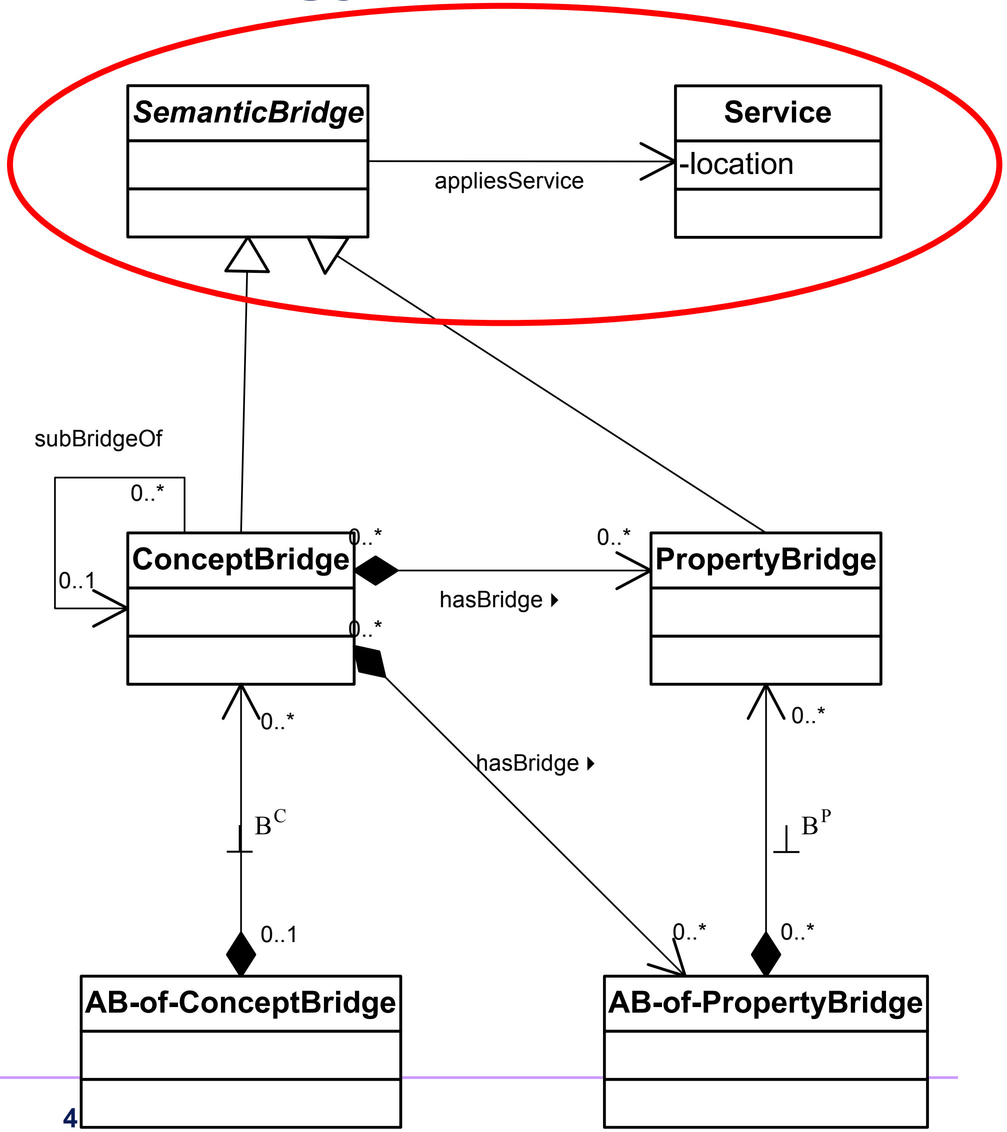
- Ontology mapping fundamentals
- Ontology mapping negotiation introdution
- Hypothesis
- Service-oriented automatic bridging
- Service-oriented negotiation
- Contributions and future work

Ontology Mapping: simple perspective



SBO - Semantic Bridge Ontology

- Taxonomy of bridges:
  - Concept Bridge
  - Property Bridge
  - Alternative Bridge
- Relation between bridges
  - subBridgeOf
  - hasBridge
- An ontology mapping specification is an instantiation of the SBO



Ontology Mapping Negotiation: Context

- Minimal or no research on the topic. None in both:
  - MeaN'2002: Meaning Negotiation WS at AAAI-02
  - MCN'2004: Meaning Coordination & Negotiation WS at ISWC-2004

- Agent and E-commerce research may be useful, but (typically):
  - One provider / Multi-consumers
  - Object of the negotiation: 1 item, as is
  - Value-oriented ($) auctions

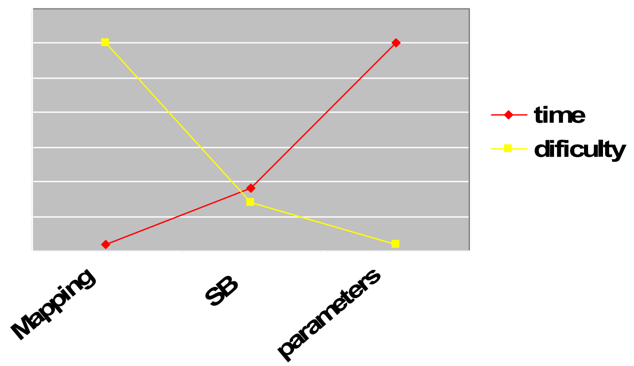
Characterization of the problem

- Negotiation
  - relaxation of the goals to be achieved by the intervenients in the negotiation, so that both achieve an acceptable contract, and as good as possible
- Intervenients
  - Cardinality and Type (ontology “owners”, mediators, facilitators)
  - Characteristics (honesty, bluffing)
- Goals
  - Object of the negotiation: mapping, semantic bridge, its parameters
  - Value of the object: correctness, relevance (both subjective)
  - Domain of the negotiation: (price, warranty, delivery, etc.)
- Relaxation mechanisms
  - What to relax: domain of the negotiation
  - How to measure relaxation efforts

Definitions

- Intervenients
  - Two ontology owners
  - Honest and non-bluffing
  - Able to derive a Mapping Document (set of Semantic Bridges)

- Goals
  - Object of the negotiation: semantic bridge
  - Value of the object: correctness + relevance
  - Domain of the negotiation: semantic bridges



Hypothesis

- **Goal/Value of the negotiation: utility function** $u(p_1, p_2, ..., p_n)$
- **Relaxation mechanisms: meta-utility function** $U(p_1, p_2, ..., p_n)$

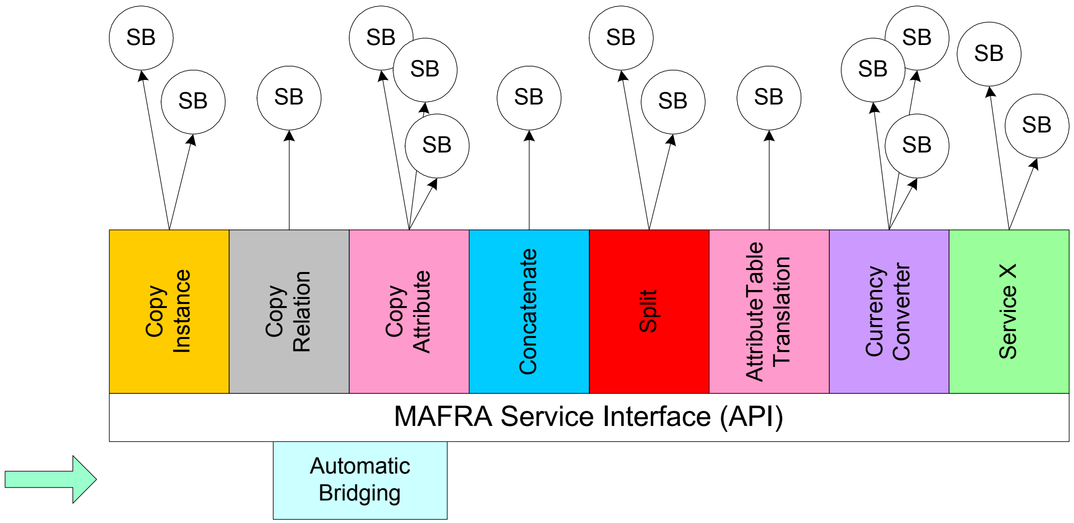
$$p_1, p_2, ..., p_n?$$

Matches

Matches represent the confidence that specific and specialized algorithms (Matchers) have, concerning the semantic similarity of two entities from two ontologies.

Metaphor

id	Source Entity	Target Entity	Matcher	Value	Justif.
m11	Individual	Woman	MOMIS-like	0,78	-
m10	Individual	Man	MOMIS-like	0,78	-
m9	Individual	Individual	MOMIS-like	0,78	-
m8	name	surname	Hyponymic	1	-
m7	name	given_name	Hyponymic	1	-
m6	spouseIn	noMarriages	Resnik-like	0,66	-
m5	name	surname	Resnik-like	0,82	-
m4	name	given_name	Resnik-like	0,82	-
m3	Individual	Woman	Resnik-like	0,86	-
m2	Individual	Man	Resnik-like	0,86	-
m1	Individual	Individual	Resnik-like	0,95	-



Parameterization

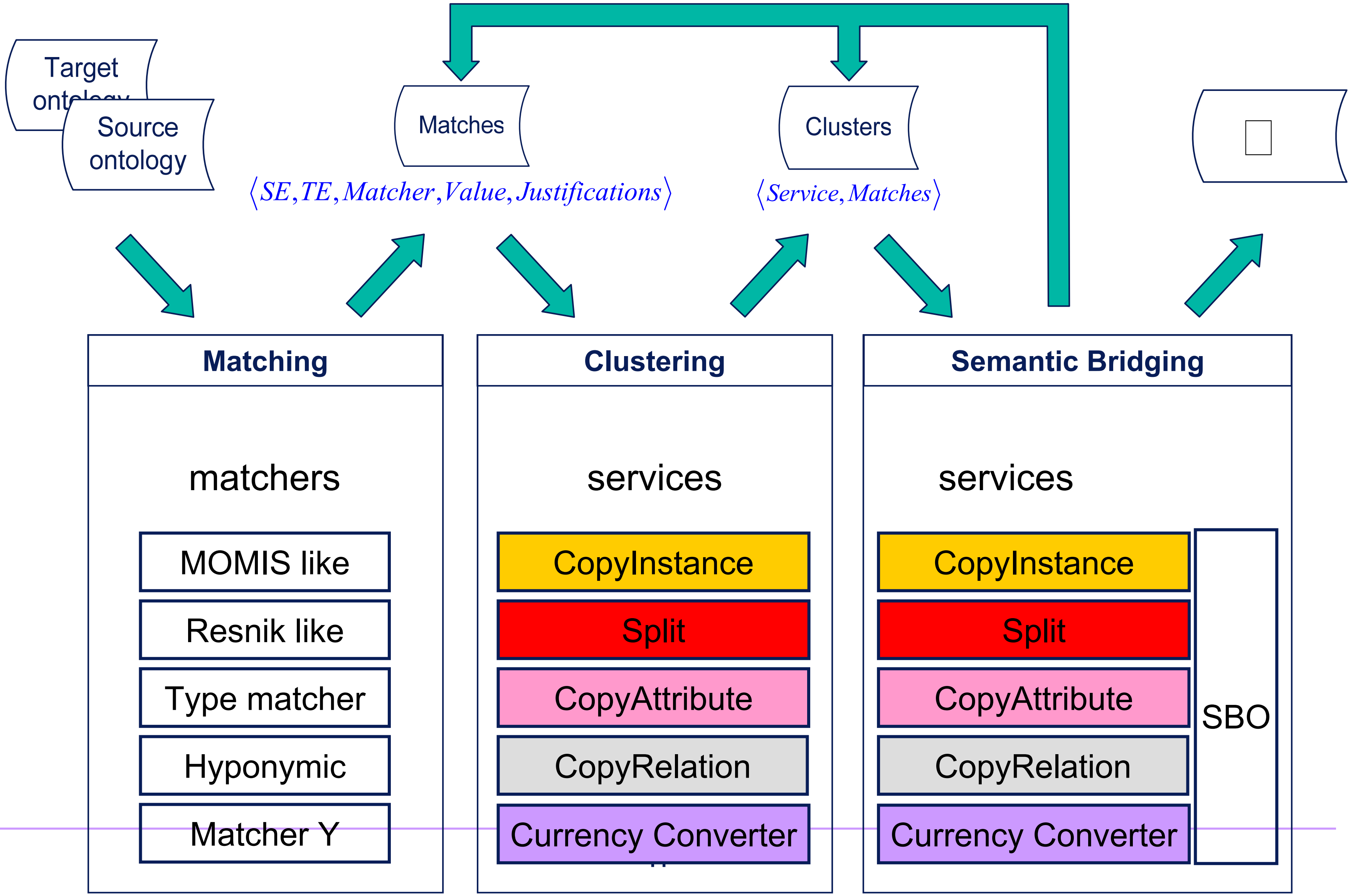
matchers

- MOMIS like
- Resnik like
- Type matcher
- Hyponymic
- Matcher Y

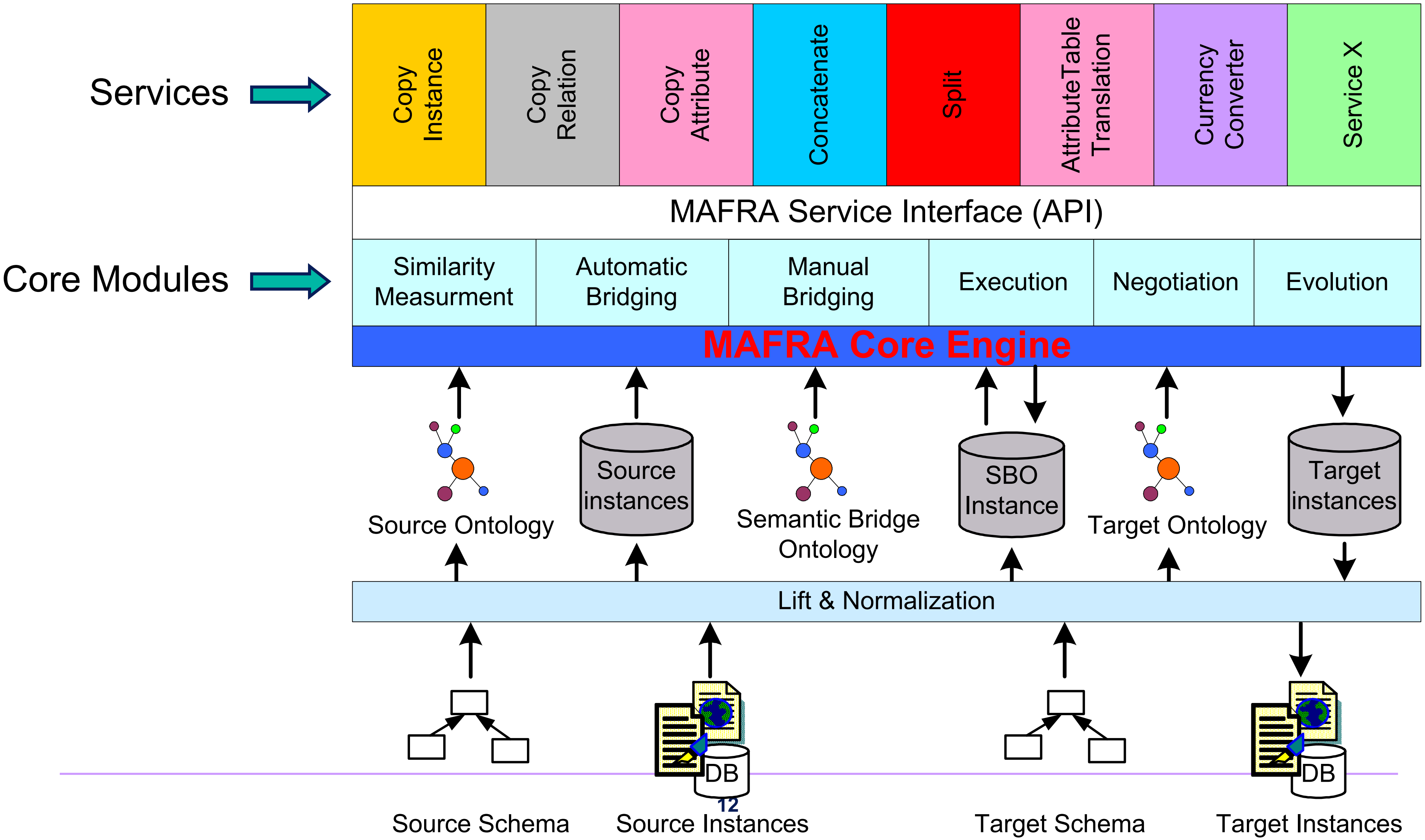
Service	Considered matches types	t_{match}	u	t_r	Extra requirements
CopyInstance	Resnik-like	0,7	u_{ci}	0,6	
	MOMIS-like	0,7			
Split	Resnik-like	0,5	u_s	0,67	
	MOMIS-like	1			
CopyAttribute	Resnik-like	0,8	u_{ca}	0,95	
	MOMIS-like	0,8			
CopyRelation	Resnik-like	0,75	u_{cr}	0,75	
	MOMIS-like	0,8			
Currency Converter	Resnik-like	$0,3<Y<0,5$	u_{cc}	0,92	Source and target attributes should be of type "currency"
	Type matcher	1			

- Combine values from matchers into an overall similarity value (**u**)
- Apply thresholds (**t_r**), determining relevance

Matchers, Clusters & Bridges



MAFRA System Architecture



Negotiation approach: basic idea

- Take advantage of the multi-dimensional service-oriented ontology architecture

- Build common consensus about similarity values proposed by Services

- Problem: How to make agents to converge to a common consensus?

Convergence process

- Let agents relax the similarity requirements (thresholds, parameters, etc.)

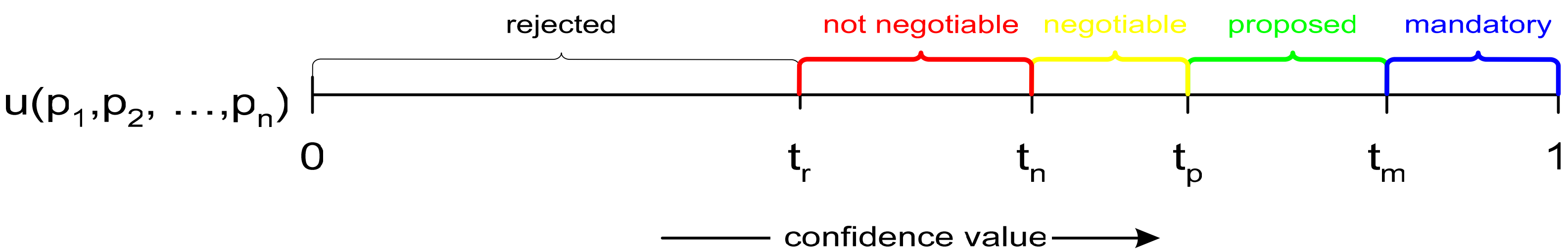
- Such that, for each agent, the sum of the similarity values associated with the consensually adopted semantic bridges is greater than without the negotiation

- Define variation functions (meta-function) upon the parameters and threshold of the utility function, determining the new value and the convergence effort

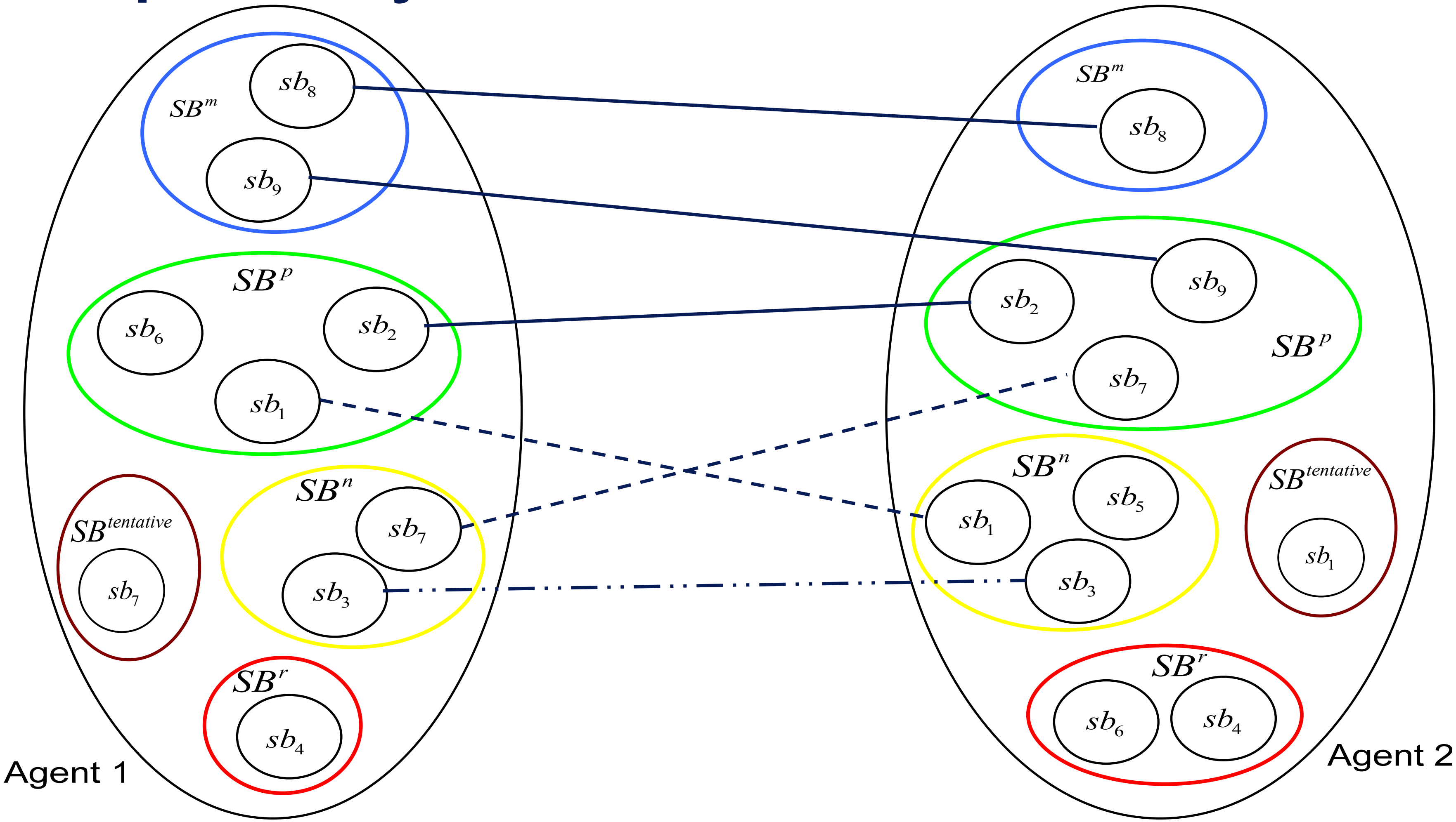
- Eventually considering preferences upon the variation of the parameters

Confidence thresholds

Service	Considered matches types	t_{match}	u	t_r	t_n	t_p	t_m	U
CopyInstance	Resnik-like	0,7	u_{ci}	0,6	0,65	0,80	0,95	U_{ci}
	MOMIS-like	0,7						
Split	Resnik-like	0,5	u_s	0,67	0,7	0,80	0,92	U_s
	MOMIS-like	1						
CopyAttribute	Resnik-like	0,8	u_{ca}	0,95	0,95	0,95	0,95	U_{ca}
	MOMIS-like	0,8						
CopyRelation	Resnik-like	0,75	u_{cr}	0,75	0,75	0,83	0,94	U_{cr}
	MOMIS-like	0,8						
Currency Converter	Resnik-like	0,3<Y<0,5	u_{cc}	0,92	0,92	0,92	0,92	U_{cc}
	Type matcher	1						



Metaphorically



Convergence effort: $e_{sb} = \mathrm{U}\left(\mathrm{p}_1, \mathrm{p}_2, ..., \mathrm{p}_\mathrm{n}\right)$

Global acceptance

$$balance = \sum c_{sb} - \sum e_t$$

$$sb \in SB^p \cup SB^m$$

$$t \in SB^{tentative}$$

< 0 - loss → Resulting document mapping is rejected → Revise

>= 0 - no loss → Resulting document mapping is accepted

Contributions and future work

- **Contributions:**
  - Characterization of the ontology mapping negotiation problem
  - Negotiation based on the utility and meta-utility functions
  - The identification of matches as parameters for these functions
  - The service-oriented negotiation process based on the categorization of semantic bridges
- **Future work**
  - Configuration and customization of the meta-utility function
  - Experiments in “real world” cases

Thanks!

Any questions?

Nuno Silva

GECAD – Knowledge Engineering and Decision Support Research Group

Polytechnic Institute of Porto - Portugal

http://www.gecad.isep.ipp.pt

Nuno.Silva@dei.isep.ipp.pt