

Intelligent Systems on the World Wide Web

5 Interoperation

Lecture Slides

Steffen Staab

Karlsruhe University

With acknowledgements to Prasenjit Mitra & Jan Jannink,
Stanford University & Jürgen Angele, Ontoprise, Karlsruhe

Outline

- Introduction
- Preliminaries
- Articulation Generation Toolkit
- Ontology Algebra
- Conclusion

Slide 2

Interoperation of Information Sources

- Compose information
- Multiple independent, heterogeneous sources
- Reliability, scalability
- Semantic heterogeneity
 - same term different semantics
 - different term same semantics

Slide 3

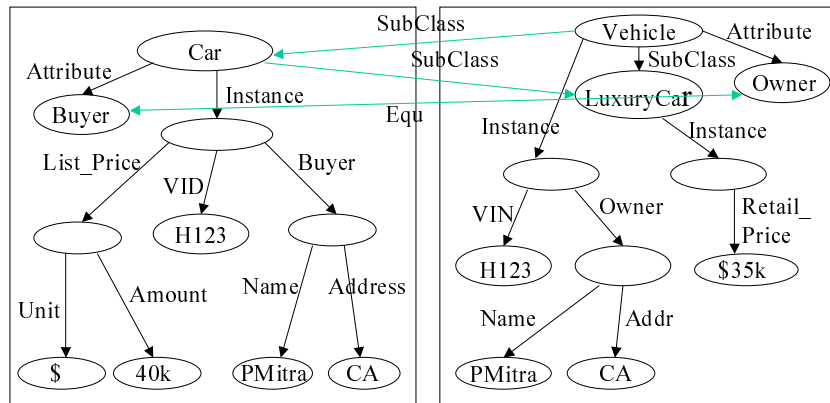
Preliminaries: Ontology

- Ontology - hierarchy of terms and specification of their properties.
- Modeled as a directed labeled graph + set of rules.
- $\text{Ont} = (V, E, R)$
 - V - set of nodes(concepts)
 - E - set of edges(properties)
 - R - set of rules involving V,E

Slide 4

Interoperation

Example



Slide 5

Interoperation

Articulation Rules



- Articulation rules - logic-based rules that relate concepts in two ontologies:
- Binary Relationships
 - (O1.Car SubClassOf O2.Vehicle)
 - (O1.Buyer Equ O2.Owner)
 - (O2.LuxuryCar SubClassOf O1.Car)

Slide 6

Interoperation

Articulation Rules (contd.)



- Horn Clauses
 - (O1.Car O1.Instance X),
(X O1.Price Y), (Y > \$30000)
=> (O2.LuxuryCar Instance X)
 - (V O2.Retail_Price P), (C O1.List_Price L),
(L Unit U), (L Amount A), (V Equ C)
=> (P Equ concat(U,A))

Slide 7

Interoperation

Contributions

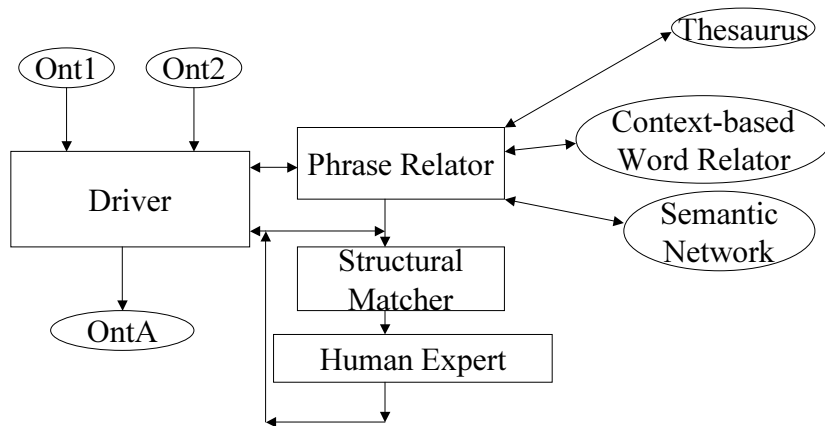


- Articulation Generation Toolkit
 - produce translation rules semi-automatically
 - a library of reusable heuristic methods
 - a GUI to display ontologies and interact with the expert
- Ontology Algebra
 - query rewriting and planning

Slide 8

Interoperation

Articulation Generator



Slide 9

Interoperation

Articulation Generation Methods

- Non-iterative Methods
 - Lexical Matcher
 - Thesaurus-based Matcher
 - Corpus-based Matcher
 - Instance-based Matcher
- Iterative Methods
 - Structural Matcher
 - Inference-based Matcher

Slide 10

Interoperation

Lexical Methods

- Preprocessing rules.
 - Expert-generated seed rules.
e.g., (O1.List_Price Equ O2.Retail_Price)
 - Context-based preprocessing directives.
e.g., (O1.UK_Govt Equ O2.US_Govt)
 - Stop-word Removal & Stemming
 - Word match (full or partial)
 - Phrase match
e.g., (O1.Ministry_Of_Defence Equ
O2.Defense_Ministry) 0.6

Slide 11

Interoperation

Thesaurus-based methods

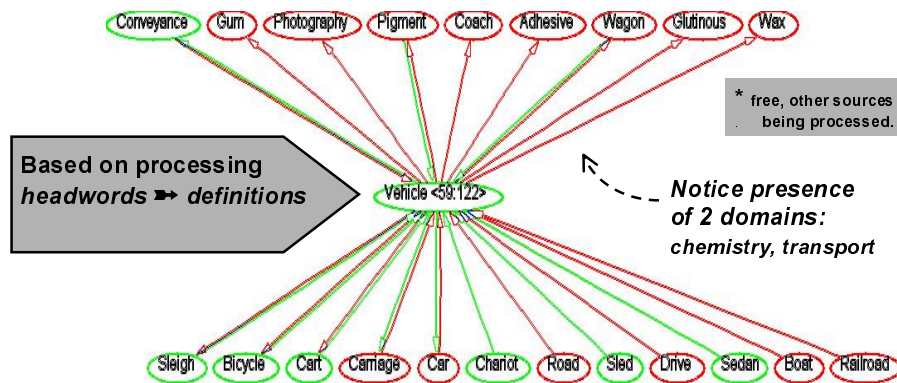
- Consult a dictionary/thesaurus to find synonyms, related words
- Generate a similarity measure or relatedness measure
 - words that have similar words in their definitions are similar
- Get more semantically meaningful relationships from WordNet (syn, hyper)

Slide 12

Interoperation

Candidate Match Repository

Term linkages automatically extracted from 1912 Webster's dictionary *



Slide 13

Interoperation

Corpus-based

- Collect a set of text documents preferably from same domain
 - search using keywords in google
- Build a context vector (1000-character neighbourhood) for each word
- Compute word-pair similarity based on the cosine of the vectors
- Use word-pair similarity to find similarity among labels of nodes/edges

Slide 14

Interoperation

Structural Methods

- Uses results of lexical match
- If x% of parent nodes match & y% of children nodes match
- Special relations (AttributeOf) match

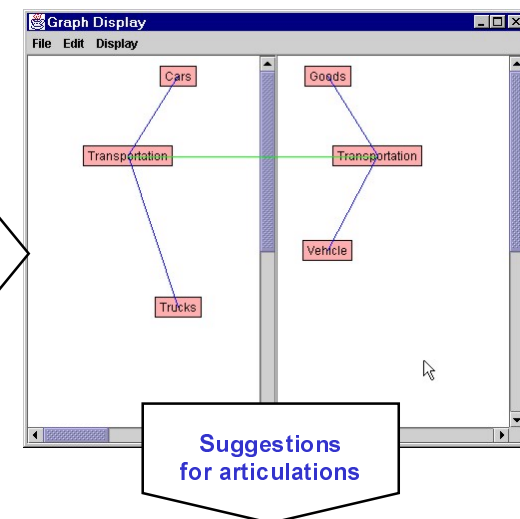
Slide 15

Interoperation

Tools to create articulations

Graph matcher for
Articulation-
creating
Expert

Vehicle
ontology



Transport
ontology

Slide 16

continue from initial point

Also suggest similar terms
for further articulation:

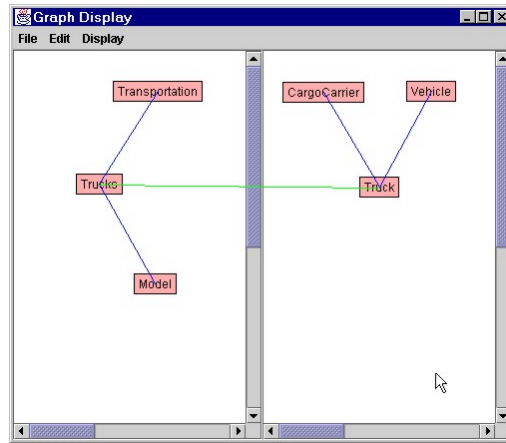
- by spelling similarity,
- by graph position
- by term match repository

Expert response:

1. **Okay**
2. **False**
3. **Irrelevant**
to this articulation

All results are recorded

Okay 's are converted into articulation rules



Slide 17

An Ontology Algebra

Operations can be composed
Operations can be rearranged
Alternate arrangements can be evaluated
Optimization is enabled
The record of past operations can be kept and reused when sources change

Slide 18

Binary Operators

A knowledge-based algebra for ontologies

Intersection		create a subset ontology keep sharable entries
Union		create a joint ontology merge entries
Difference		create a distinct ontology remove shared entries

articulation rules between them.

Slide 19

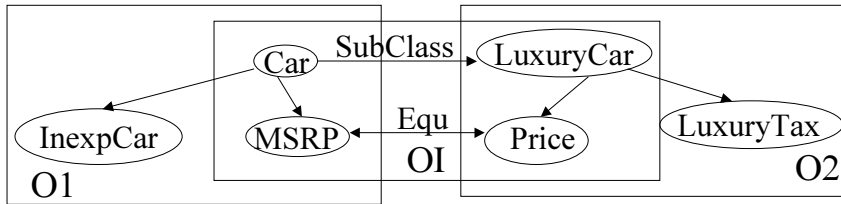
Intersection: Definition

- $O1 = (V1, E1, R1)$ $O2 = (V2, E2, R2)$
- $OI = (O1 \text{ Int}_{\text{ArtGen}} O2) = (VI, EI, RI)$
- $Arules = \text{ArtGen}(O1, O2)$
- $VI = \text{Nodes}(Arules)$
- $EI = \text{Edges}(Arules) + \text{Edges}(E1, VI.V1) + \text{Edges}(E2, VI.V2)$
- $RI = Arules + \text{Rules}(R1, VI.V1) + \text{Rules}(R2, VI.V2)$

Slide 20

Interoperation

Intersection: Example



- $ARules = \{ (O1.Car \text{ SubClass } O2.LuxuryCar), (O1.MSRP \text{ Equ } O2.Price) \}$
- $NI = (O1.Car, O1.MSRP, O2.LuxuryCar, O2.Price)$
- $EI = Edges(ARules) + \{ (O1.Car \text{ Attribute } O1.MSRP), (O2.LuxuryCar \text{ Attribute } Price) \}$

Slide 21

Interoperation

Intersection: Properties



- Commutative?

$$OI_{12} = (O1 \text{ Int}_{ArtGen} O2) = (O2 \text{ Int}_{ArtGen} O1) = OI_{21}$$

$$VI_{12} = VI_{21}, EI_{12} = EI_{21}, RI_{12} = RI_{21}$$

$$ARules_{12} = ARules_{21}$$

$$ArtGen(O1, O2) = ArtGen(O2, O1)$$

- Int_{ArtGen} is commutative iff $ArtGen$ is commutative

Slide 22

Interoperation

Semantically Commutative



Example:

- $ArtGen(O1, O2) : (O1.Car \text{ SubClassOf } O2.Vehicle)$
- $ArtGen(O2, O1) : (O2.Vehicle \text{ SuperClassOf } O1.Car)$
- Defn: $ArtGen$ is Semantically Commutative iff $ArtGen(O1, O2) \Leftrightarrow ArtGen(O2, O1)$

Operands to intersection can be rearranged if $ArtGen$ is semantically commutative

Slide 23

Interoperation

Associativity Example



- Example:
- $ArtRules(O1, O3) : (O1.Car \text{ SubClassOf } O3.Vehicle)$
- $ArtRules(O2, O3) : (O2.Truck \text{ SubClassOf } O3.Vehicle)$
- $ArtRules(O1, O2) : \text{null}$
- Intersection is not associative!

Slide 24

Associativity



- $(O1.O2).O3 = O1.(O2.O3)$
iff ArtGen is consistent and transitively connective
- consistent - given two nodes it generates the same relationship irrespective of relationships between other nodes

Slide 25

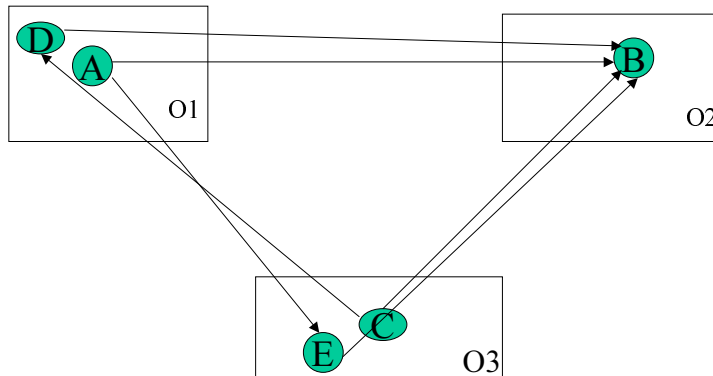
Associativity (contd.)



- ArtGen relates A and B:
 $(A \text{ Rel } B) = (A \text{ R1 } B) \text{ or } (B \text{ R2 } A)$
- Transitively connective:
If ArtGen generates $(A \text{ Rel } B)$, $(B \text{ Rel } C)$
then it also generates $(A \text{ Rel } C')$
where $A \in O1$, $B \in O2$, $C, C' \in O3$

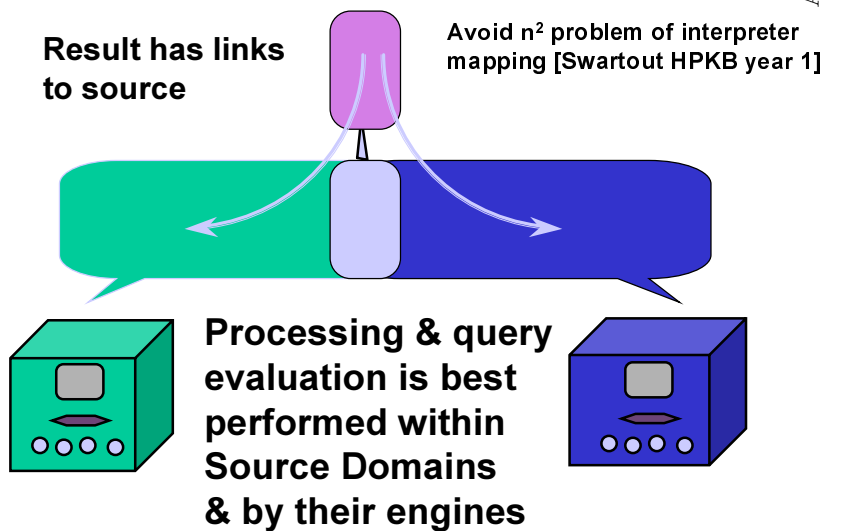
Slide 26

Transitive Connectivity



Slide 27

Exploiting the result



Slide 28

Interoperation



SKEIN Algorithms – Use Structure and corpora to derive articulation rules

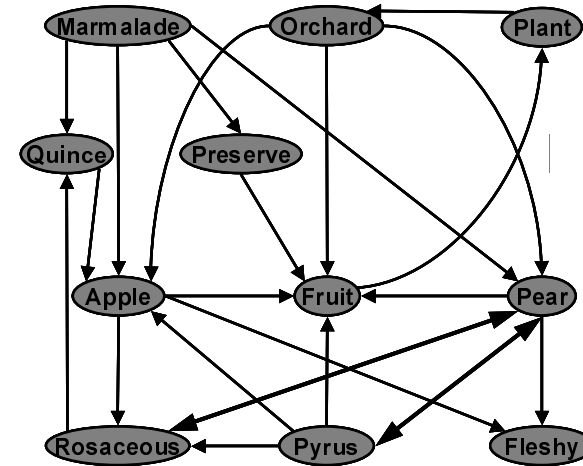
- PageRank
- Relative Arc Importance
- ArcRank
- All Pairs Similarity
- Qualitative Properties

Slide 29

Interoperation



PageRank



Slide 30

Interoperation



Intuitive Scenario of Page Rank

- Surfing the web
- Following links only forward
- Choosing with same probability one of the outgoing links of a page
- Sometimes getting bored and starting an arbitrary new page
- The page rank of a page P is given by the stationary probability of being at P

Slide 31

Interoperation

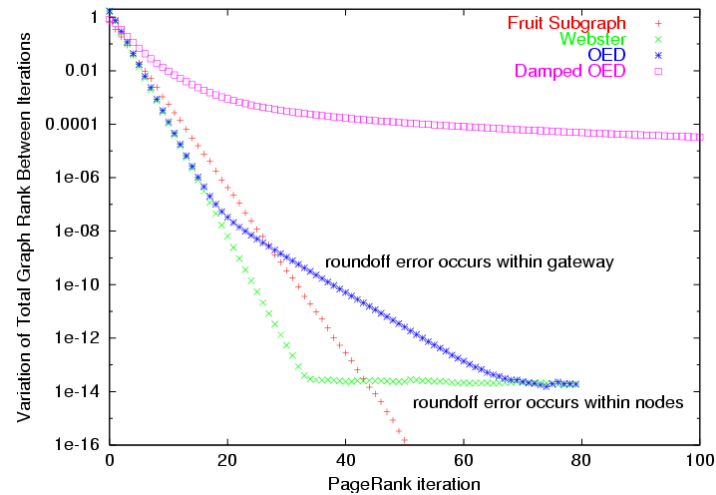


- $PR(P) = (1-d) + d (PR(In_1)/\#In_1 + \dots + PR(In_n)/\#In_n)$
- d is a damping factor („getting bored“)
- In_i are the pages referring to P
- $PR(In_i)$ is the probability of being at In_i
- $\# In_i$ is the number of outgoing links of In_i

Slide 32

Interoperation

Convergence Results



Slide 33

Interoperation

Relative Arc Importance

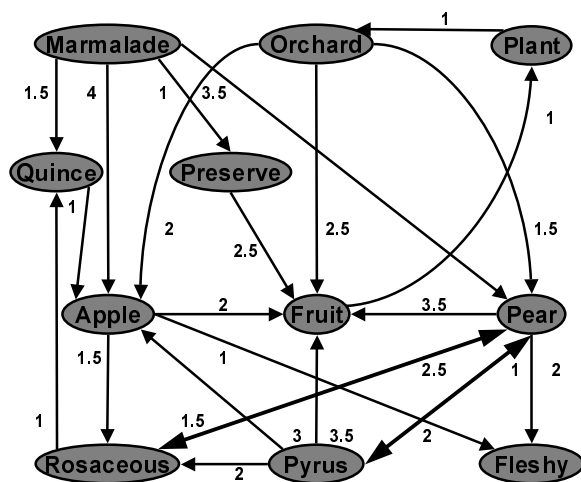
- PageRank limitations
 - **node oriented**
 - **ranks highly words with low semantic contribution**
 - conjunctions [And](#)
 - articles [The](#)
 - prepositions [to](#)
 - pronouns [it](#)
- Relative arc importance
 - **arc oriented**
 - **contribution of source rank P_s to target rank P_t**

$$v_{s,t} = \sum_{e_{s,t}} \frac{p_s / |a_s|}{p_t}$$

Slide 34

Interoperation

ArcRank



Arc Rank

Slide 35

Interoperation

All Pairs Similarity

Usage Arc
Importance Vector

Marmalade
Orchard
Pyrus
Quince

→ Apple →

Definitional Arc
Importance Vector

Fruit
Fleshy
Rosaceous

$$sim_{apple, pear} = (uv_{apple} \cdot uv_{pear} + dv_{apple} \cdot dv_{pear}) / 2$$

Marmalade
Orchard
Pyrus
Rosaceous

→ Pear →

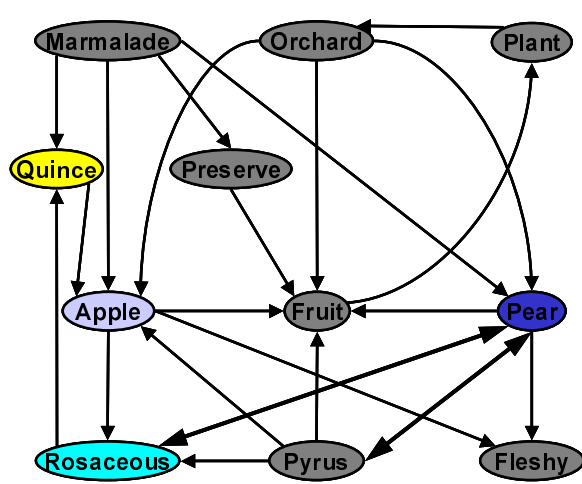
Fruit
Fleshy
Rosaceous
Pyrus

Slide 36

Interoperation

All Pairs Similarity

AIFB



```

1 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 1

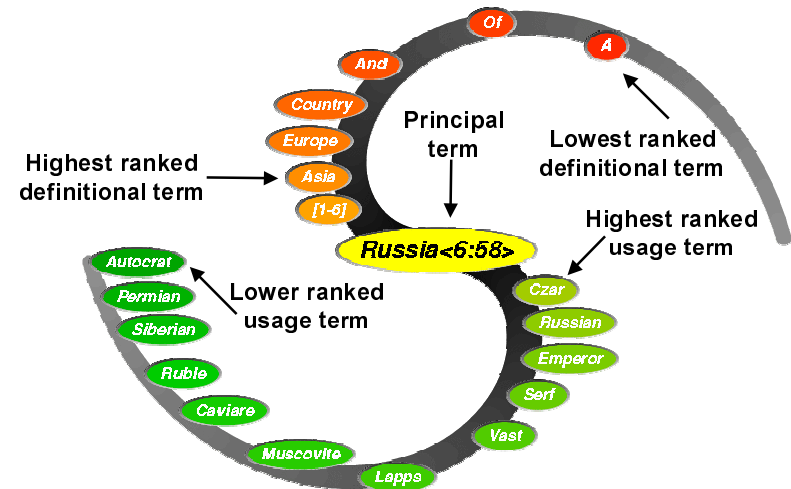
```

Slide 37

Interoperation

New Visualization GUI

AIFB

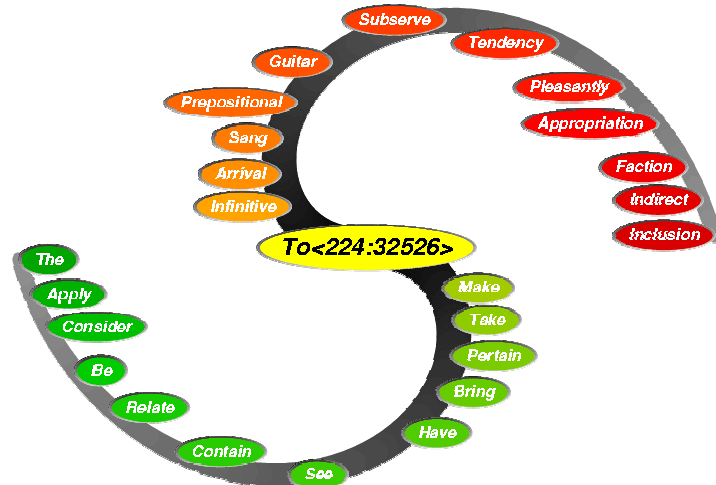


Slide 38

Interoperation

Example (Stop word)

AIFB

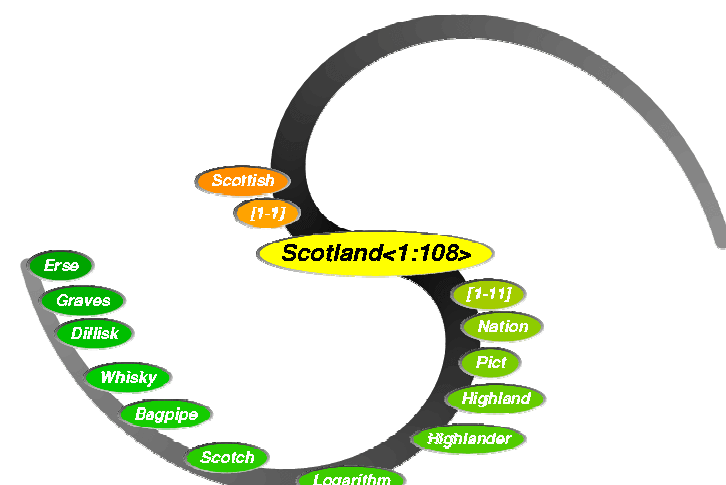


Slide 39

Interoperation

Example (Proper Noun)

AIFB



Slide 40

Qualitative Properties

- First trials using nexus to support interoperation
 - found 70% of matches contained in a hand-coded articulation
 - no false positives found
- Quality survey of ArcRank algorithm in nexus
 - completed, results being compiled
 - <http://skeptic.stanford.edu/cgi-bin/survey.cgi>
- Justification for using web algorithms on nexus
 - New survey of 200M web pages shows identical macroscopic structure [Rajagopalan]

Slide 41

Related Work

- Tsimmis
- Garlic
- Infomaster
- Information Manifold
- Clio
- LSD

Slide 42

Conclusion

- Introduced an Ontology Interoperation System.
- Interoperation based on articulations that bridge the semantic gap.
- Graph-oriented data model, logical rules
- Founded on Ontology Algebra

Slide 43