HCS

Universiteit van Amsterdam

MultimediaN

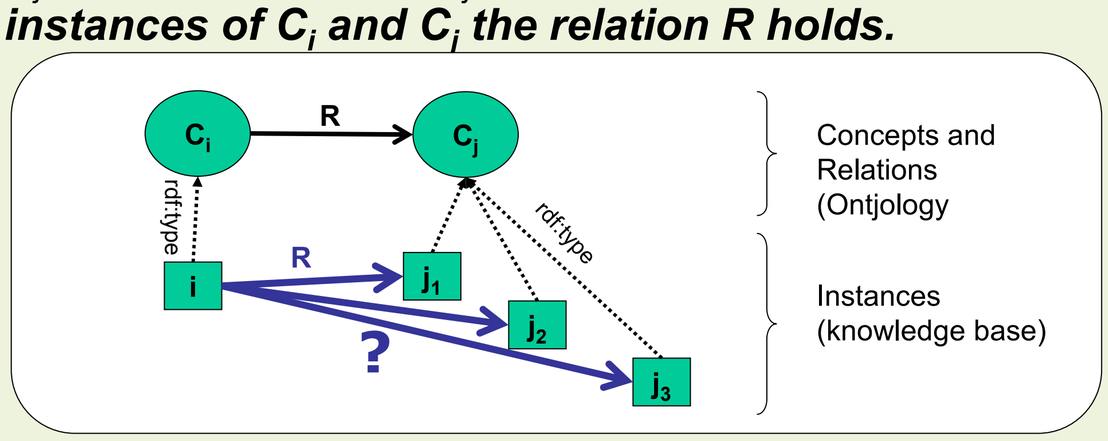
A Redundancy-based Method for Relation Instantiation from the Web

Viktor de Boer, Maarten van Someren and Bob J. Wielinga
Human-Computer Laboratories, Informatics Institute, Universiteit van Amsterdam {vdeboer, maarten, wielinga}@science.uva.nl

Task: Relation Instantiation

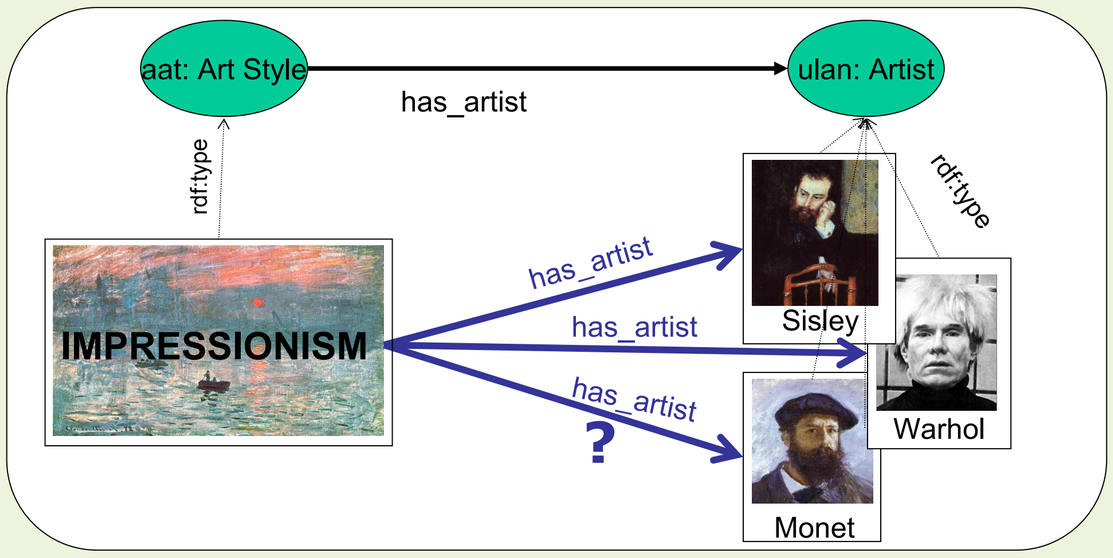
Subtask of Ontology Population:

For a domain ontology with instantiated concepts C_i and C_j and a relation $R(C_i,C_j)$: ***Identify between which instances of C_i and C_j the relation R holds.***



Domain: Cultural Heritage

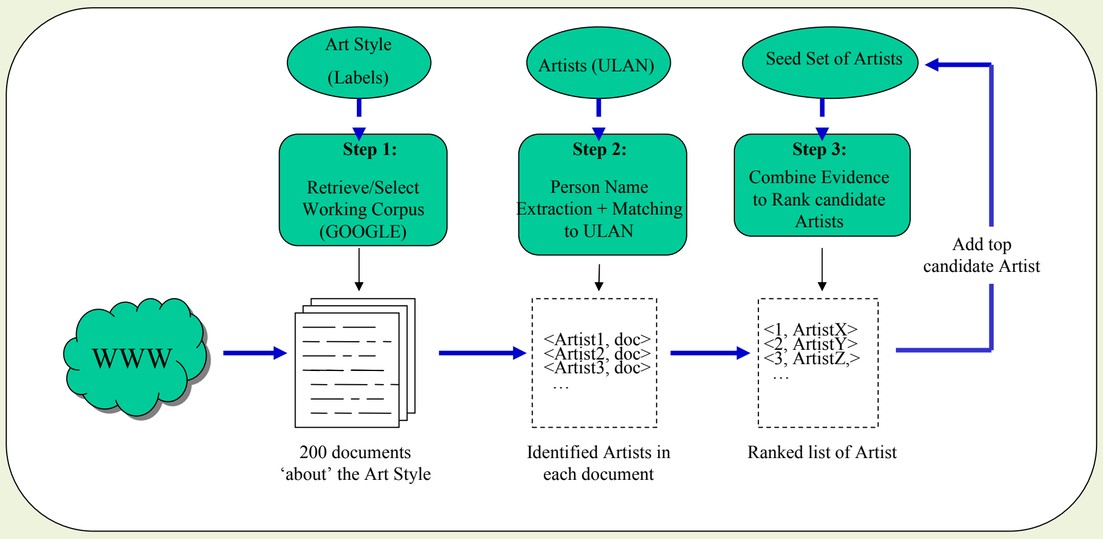
Cultural Heritage Ontology including the Art and Architecture Thesaurus (AAT) and Unified List of Artist Names (ULAN)



Domain-specific Task: ***Which Art Style is represented by which Artist?***

Approach: Use Redundancy

Exploit the observation that information is redundantly available on the WWW: use **coarse-grained** general methods to extract information from a **large number** of sources and **combine** the evidence



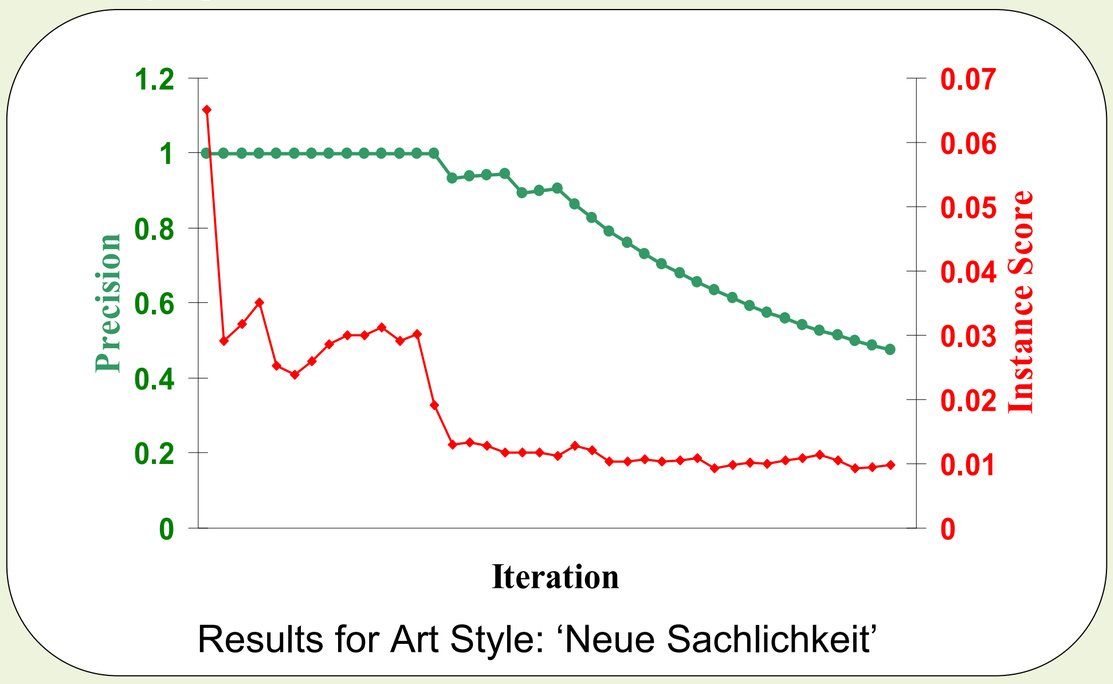
1.Retrieve docs from WWW about Art Style (Google)
2.Identify ULAN Artists in the documents (PNE/ Matching)
3.Combine the evidence for each candidate to end up with a ranking. Add top candidate relation. Redo step 3.

Ranking uses simple co-occurrence with seed set:

- Every doc gets a document score
  - (Artists in Doc and in Seed Set / Artist in Doc)
- Every artists gets an instance score (evidence)
  - Normalized Sum of Document Scores

Experiments

- 9 modern art styles, 3 artists in seed set, 1000 pages each
- Manual evaluation using fixed set of authoritative web pages



Results for Art Style: 'Neue Sachlichkeit'

- Iteration Threshold Parameters:
  - *Drop Factor (DF)*: Stop if evidence drops
  - *Max:* Stop after maximum no.of iterations

	Max							
	10		20		30		40	
DF	prec	ex	prec	ex	prec	ex	prec	ex
0	0.856	77	0.806	145	0.722	195	0.65	234
0.1	0.856	77	0.806	145	0.721	193	0.648	228
0.2	0.856	77	0.799	137	0.776	179	0.746	197
0.3	0.865	73	0.842	117	0.830	138	0.81	144
0.4	0.857	62	0.834	96	0.826	114	0.824	120
0.5	0.902	55	0.878	86	0.868	103	0.866	109
0.6	0.924	46	0.896	67	0.882	81	0.88	87

Average precision and total number of correct extractions for 9 art styles for different values of *DF* and *Max*

Conclusions

- General method based on redundancy and simple Co-occurrence measures for Relation Instantiation
- Tested in CHD domain on extracting Art Style-Artist relation instances: F1 Measures and Precision satisfactory
- Further Research:
  - Other domains / relations
  - Use more ontological information
  - Combine information from different resources

Contact

Email:
vdeboer@science.uva.nl

Eculture Demo:
http://eculture.multimedian.nl/demo/search