Abstract:

Associations, which are one of the key ingredients of human intelligence and thinking, are not easily accessible to the Semantic Web community. High quality RDF datasets of this kind are missing.

In this work we generate such a dataset by transforming 788 K free-text associations of the Edinburgh Associative Thesaurus (EAT) into RDF.

Furthermore, we provide a verified mapping of strong textual associations from EAT to DBpedia Entities with the help of a semi-automatic mapping approach.

Both generated datasets are made publicly available and can be used as a benchmark for cross-type link prediction and pattern learning.

Edinburgh Associative Thesaurus (EAT) as RDF and DBpedia Mapping

Jörn Hees, Rouven Bauer, Joachim Folz, Damian Borth, Andreas Dengel

15

EAT: G. Kiss, C. Armstrong, R. Milroy, J. Piper (1973)

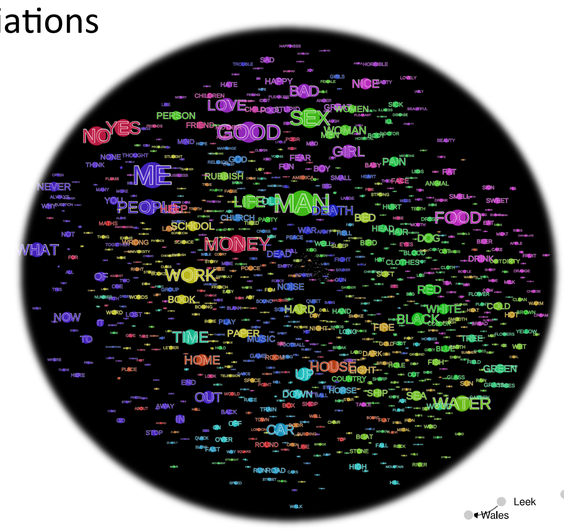
- Association corpus
  - For each stimulus asked 100 people for a response
  - Strong responses became stimuli of next round
- 790 K raw associations (free text)
- Graph: (|V| = 23 K, |E| = 325 K)
  - 5000 strong associations (>19x)
  - 167.4 K raw associations

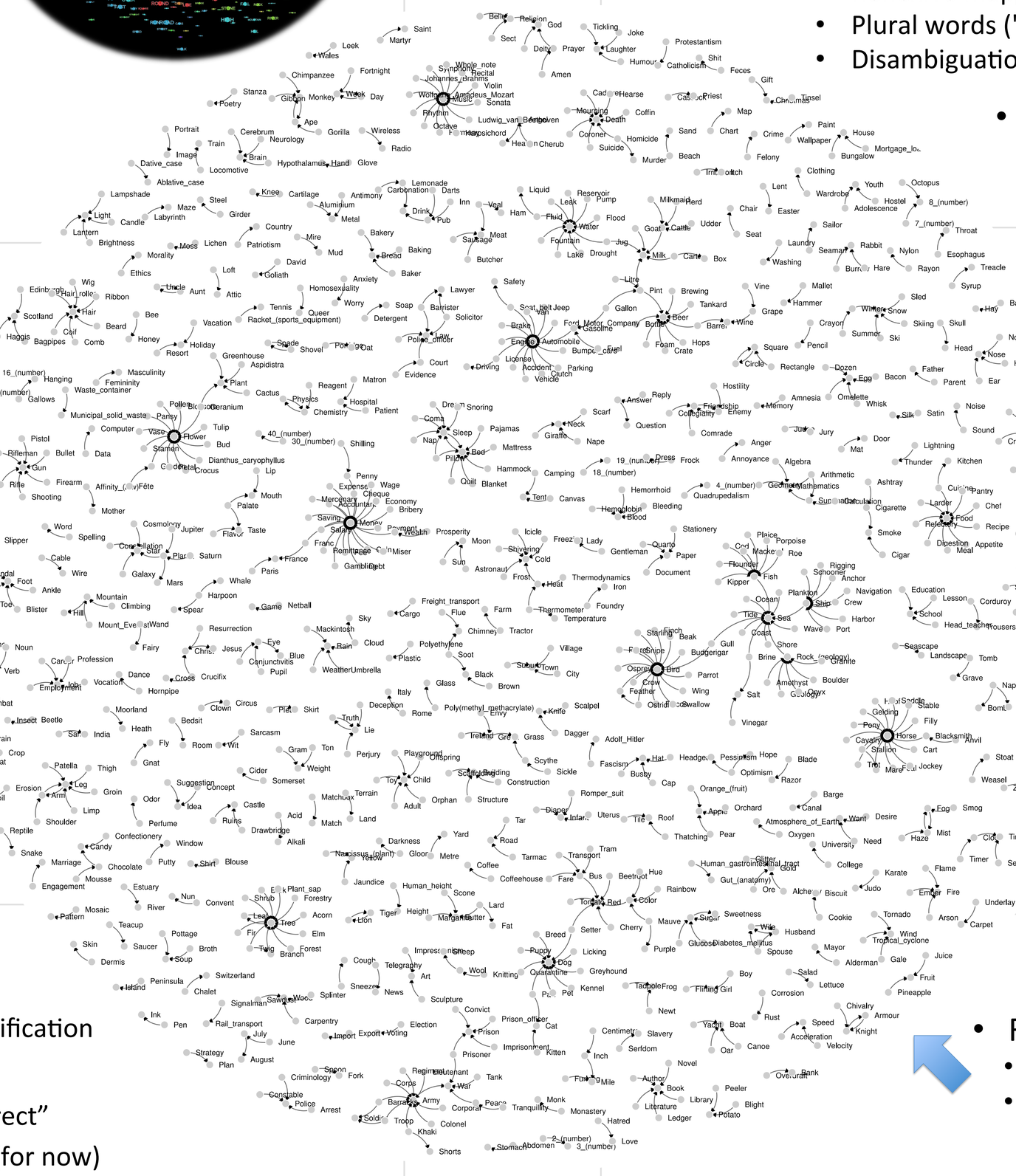
Example Associations:

dog		**cat**	
cat	57	dog	49
collar	5	mouse	8
bark	3	black	4
leg	2	mat	3
		tom	2

man		**woman**	
woman	66	man	59
strong	3	sex	5
human	2	girl	5
hole	2	female	3
boy	2	child	2

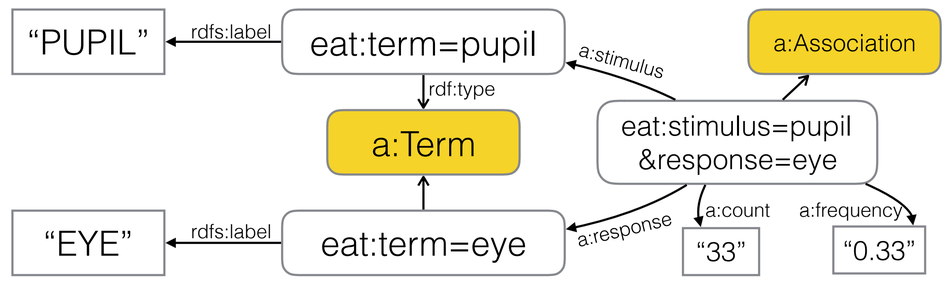
bank		**money**	
money	25	cash	7
book	7	bags	7
account	7	pounds	6
manager	6	gold	6
clerk	5	wealth	5





EAT as RDF:

- 1.6 M triples



Automatic Mapping Approach to DBpedia:

- Lookup + Scoring on Wikipedia
  - Scoring:
    - Composite phrases ("port" - "wine")
    - Word types non-nouns ("unbound" - "free")
    - Reflexive mappings ("child" - "children”)
    - Plural words ("thumbs" - "fingers")
    - Disambiguation pages ("pod" - "pea")
  - Results:
    - 1066 mapping candidates
    - ➢ Manual verification

Resulting Top Response Nodes

Response	Count	Response	Count
dbr:Money	19	dbr:Water	9
dbr:Bird	15	dbr:Army	8
dbr:Horse	14	dbr:Beer	8
dbr:Automobile	13	dbr:Death	7
dbr:Flower	12	dbr:Fish	7
dbr:Music	12	dbr:Bed	7
dbr:Tree	11	dbr:Ship	7
dbr:Sea	11	dbr:Red	6
dbr:Dog	9	dbr:Gun	6
dbr:Food	9	dbr:Hair	6

Node Degrees in Dbpedia

Vanilla:

Node	Degree
dbr:Animal	237855
dbr:Insect	118589
dbr:France	94826
dbr:India	85386
dbr:Plant	79062
dbr:Italy	55966
dbr:Village	54082
dbr:Beetle	43739
dbr:Scotland	27607
dbr:Bird	25933
dbr:Switzerland	19874
dbr:City	18030
dbr:Paris	17362
dbr:Wales	14605
dbr:Town	13301
dbr:Ireland	11340
dbr:Rome	10344
dbr:Fly	10299
dbr:Mayor	9812
dbr:Reptile	9595

Extended (Wikilinks+):

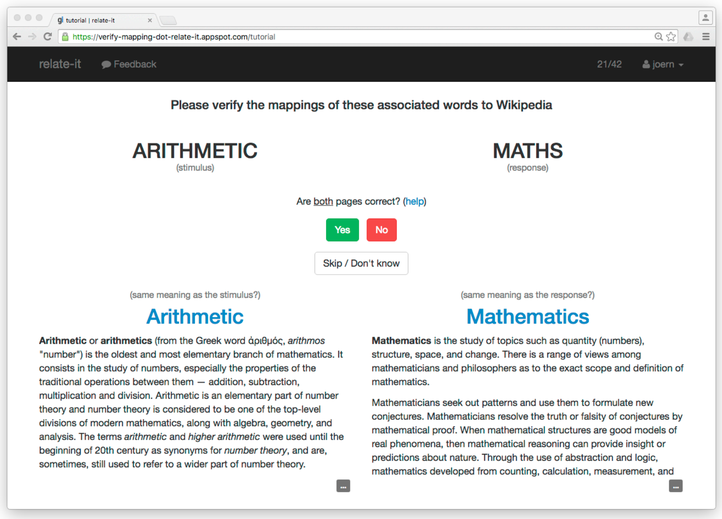
Node	Degree
dbr:Animal	445324
dbr:Village	344264
dbr:Insect	239032
dbr:France	234700
dbr:India	196686
dbr:Plant	149369
dbr:Italy	143942
dbr:Beetle	85994
dbr:Town	83109
dbr:Scotland	73312
dbr:Bird	66504
dbr:Switzerland	61214
dbr:Paris	53008
dbr:City	50332
dbr:Bird	40592
dbr:Ireland	38643
dbr:Rome	38611
dbr:Wales	38532
dbr:School	32824
dbr:Novel	32193

Mapping Verification:

- Web App
  - Allows quick manual verification
- Test group (13 people)
  - Valid: 3 people say “correct”
  - Invalid: Everything else (for now)

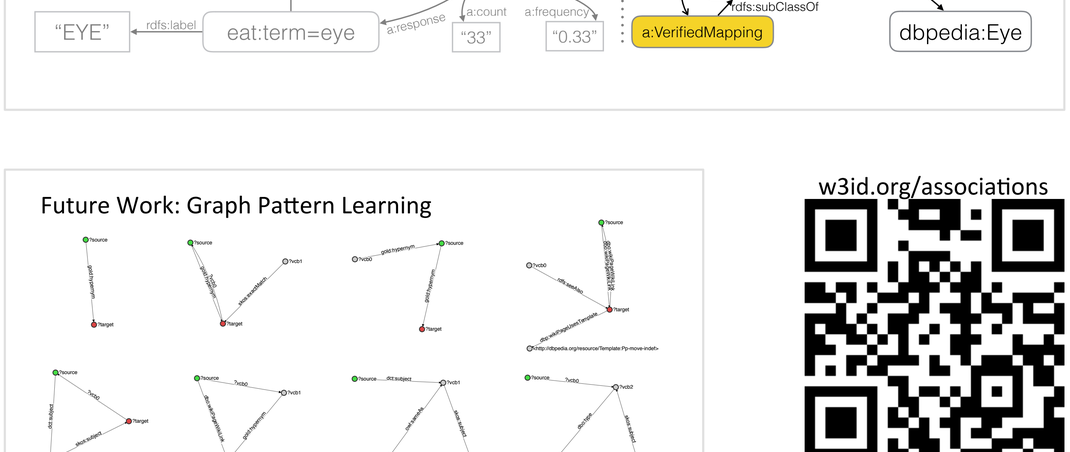
➢ 790 valid semantic associations

- 25.5 K raw associations



Semantic Associations:

- Results:
  - 727 unique associations
  - between DBpedia Entities



Future Work: Graph Pattern Learning

w3id.org/associations

joern.hees@dfki.de

German Research Center for Artificial Intelligence

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN