Knowledge Patterns as one means to overcome ontology design bottlenecks

Aldo Gangemi

aldo.gangemi@cnr.it, @lipn.univ-paris13.fr

RCLN, LIPN Université Paris 13, UMR CNRS, Sorbonne Cité

Semantic Technology Lab, Institute for Cognitive Sciences, CNR, Rome, Italy

Work described jointly with STLab people in the last years:

Valentina Presutti, Francesco Draicchio, Andrea Nuzzolese, Diego Reforgiato

Alessandro Adamou, Eva Blomqvist, Enrico Daga, Alfio Gliozzo, Alberto Musetti

Partly based on iSemantics 2013 keynote slides, available on Slideshare:

http://www.slideshare.net/gangemi/isemantics-keynote

My arguments

Semantic technologies only sparsely address real semantic phenomena

My arguments

Semantic technologies only sparsely address real semantic phenomena

Semantic Framing is not explicit

My arguments

Semantic technologies only sparsely address real semantic phenomena

Semantic Framing is not explicit

OE bottlenecks are the effect of that

My arguments

Semantic technologies only sparsely address real semantic phenomena

Semantic Framing is not explicit

OE bottlenecks are the effect of that

We need an entity-centric, frame-oriented data science to ensure *relevance*

Semantic technologies have progressed significantly, but they still miss most relevant semantic phenomena in social life

How to *synthetically* tell what a text is about, in the open domain, i.e. without specific training?

What is its core meaning, emotional content, position in the evolution of its topic, etc.?

How to *analytically* mashup data in *relevant* ways, in the open domain, i.e. without someone putting the necessary intelligence within?

Examples

- Data integration interpretation without designers (risk of *correlation fallacy*)
- Opportunistic reasoning: travel planning, financial opportunities, team building, etc.
- Smart text summarization
- Opinion mining on the right spots
- Domain dynamics: science evolution, scholar changes, market dynamics, ...

Examples

- So, very good opportunity for ontologists you say?
- Ehm, not really, or not yet in the large

Even in basic OE

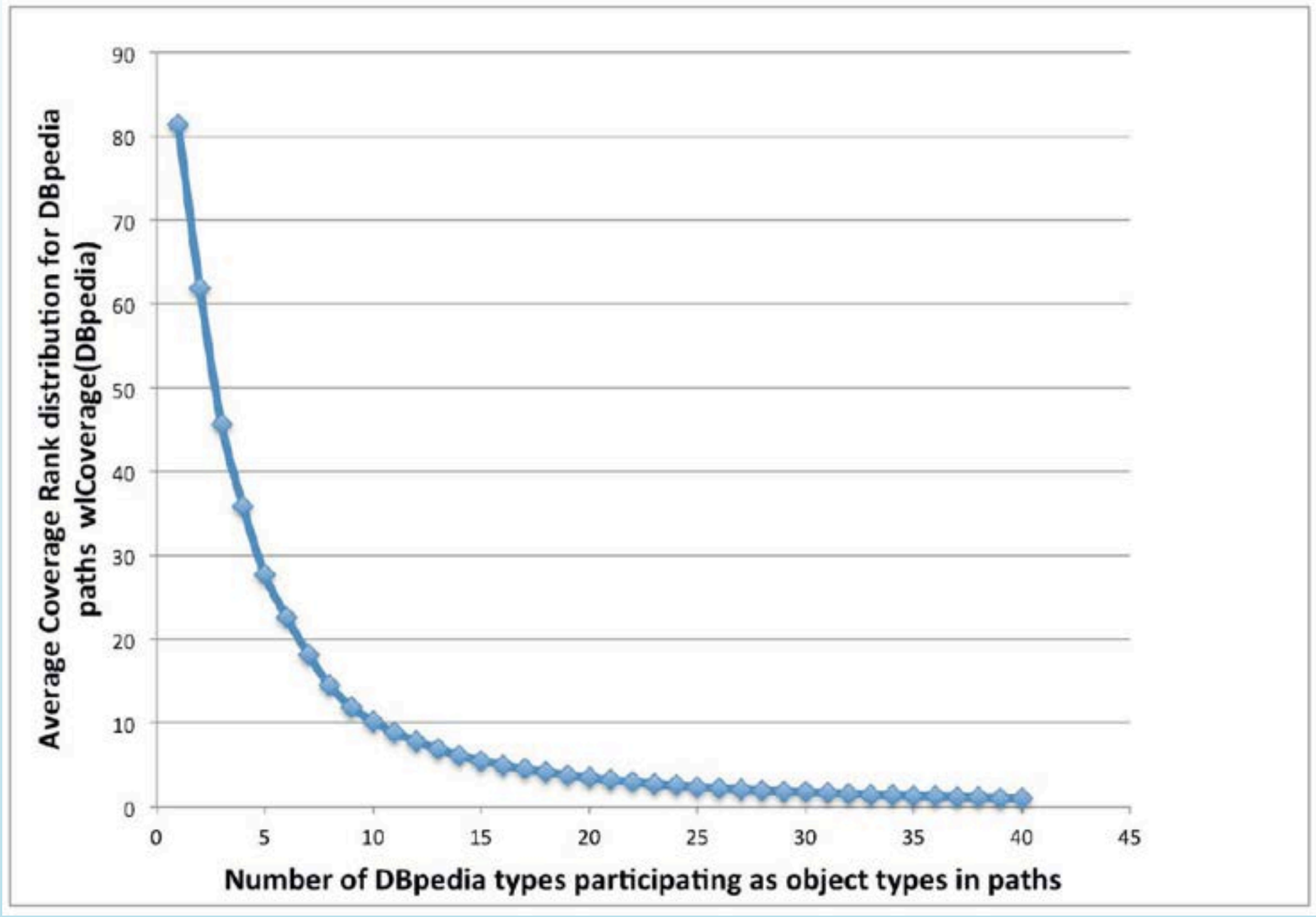
- Sociological issues with "ontologies", with giving out precious know-how, or because experts are not accustomed to reverse-engineer their own conceptualization, or because of strong competition with DB design
- I tend to reverse the argument line; we need to grasp semantics where it is, and to target the most natural way of expressing it
- **Entity-centric, frame-based computing**
- **Empirical science methods**
- **Web robustness against incompleteness and errors**

Human, social knowledge management is not exempt from *framing*: it is modulated by frames, metaphors, and stories that make something relevant through neural activation patterns

Cf. my iSemantics 2013 keynote slides, available on Slideshare:
http://www.slideshare.net/gangemi/isemantics-keynote

Ex.: *k*-means clustering on Wikipedia link popularity

Sample distribution of pathPopularity for DBpedia link paths. The y-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity

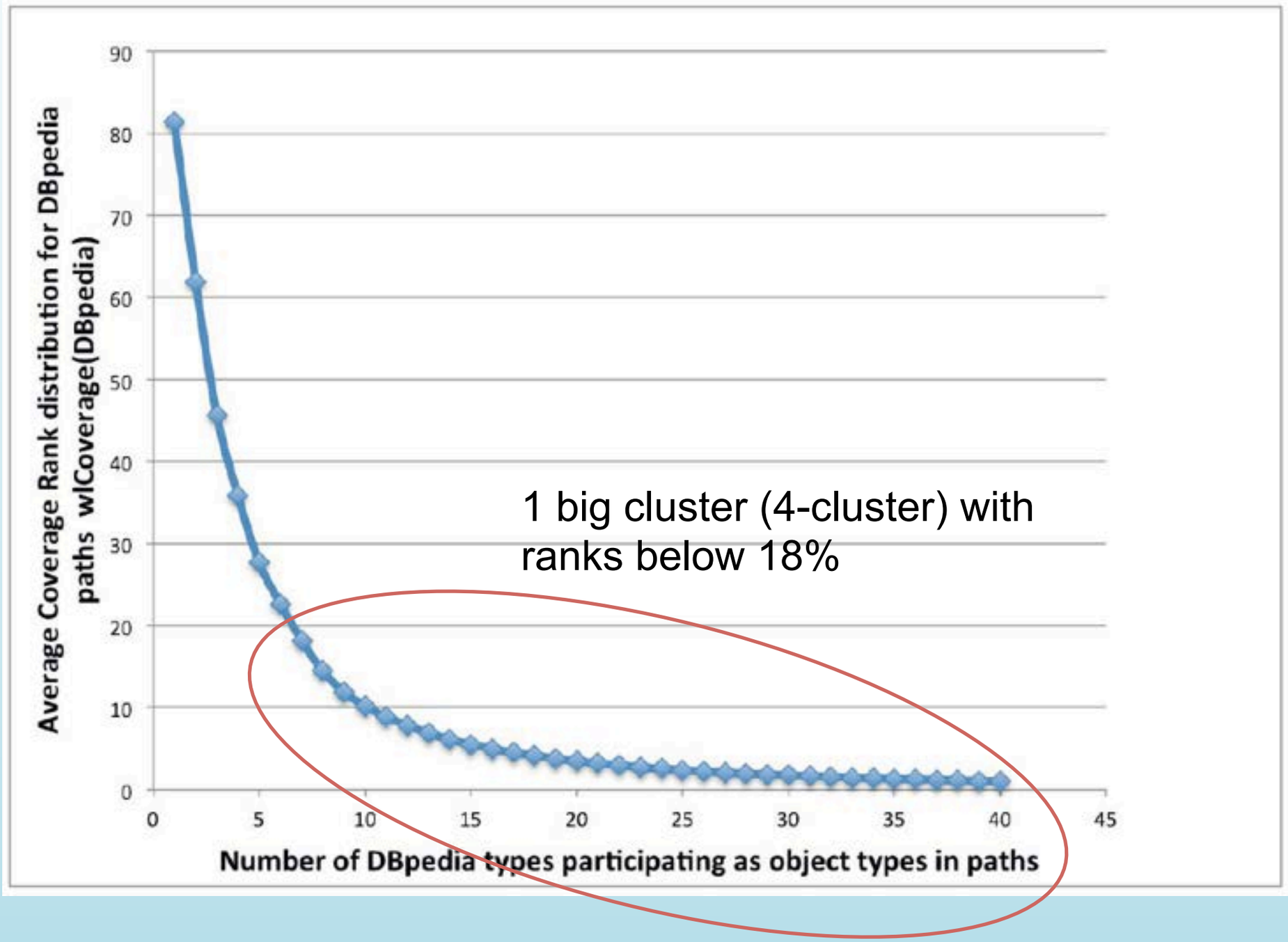


Encyclopedic Knowledge Patterns from Wikipedia Wikilinks (@ISWC2011)

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini: Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Ex.: *k*-means clustering on Wikipedia link popularity

Sample distribution of pathPopularity for DBpedia link paths. The y-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity

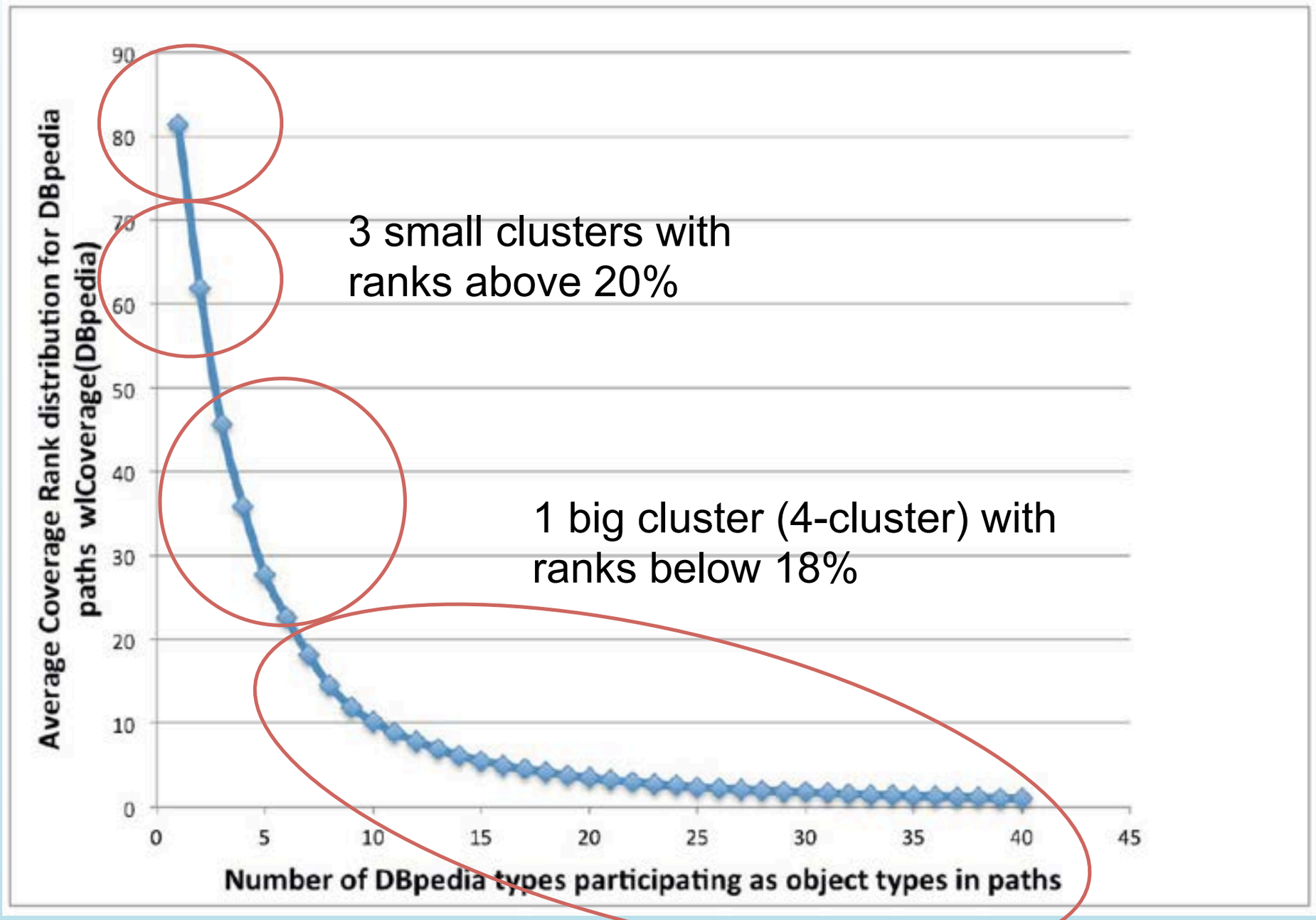


Encyclopedic Knowledge Patterns from Wikipedia Wikilinks (@ISWC2011)

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini: Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Ex.: *k*-means clustering on Wikipedia link popularity

Sample distribution of pathPopularity for DBpedia link paths. The y-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity

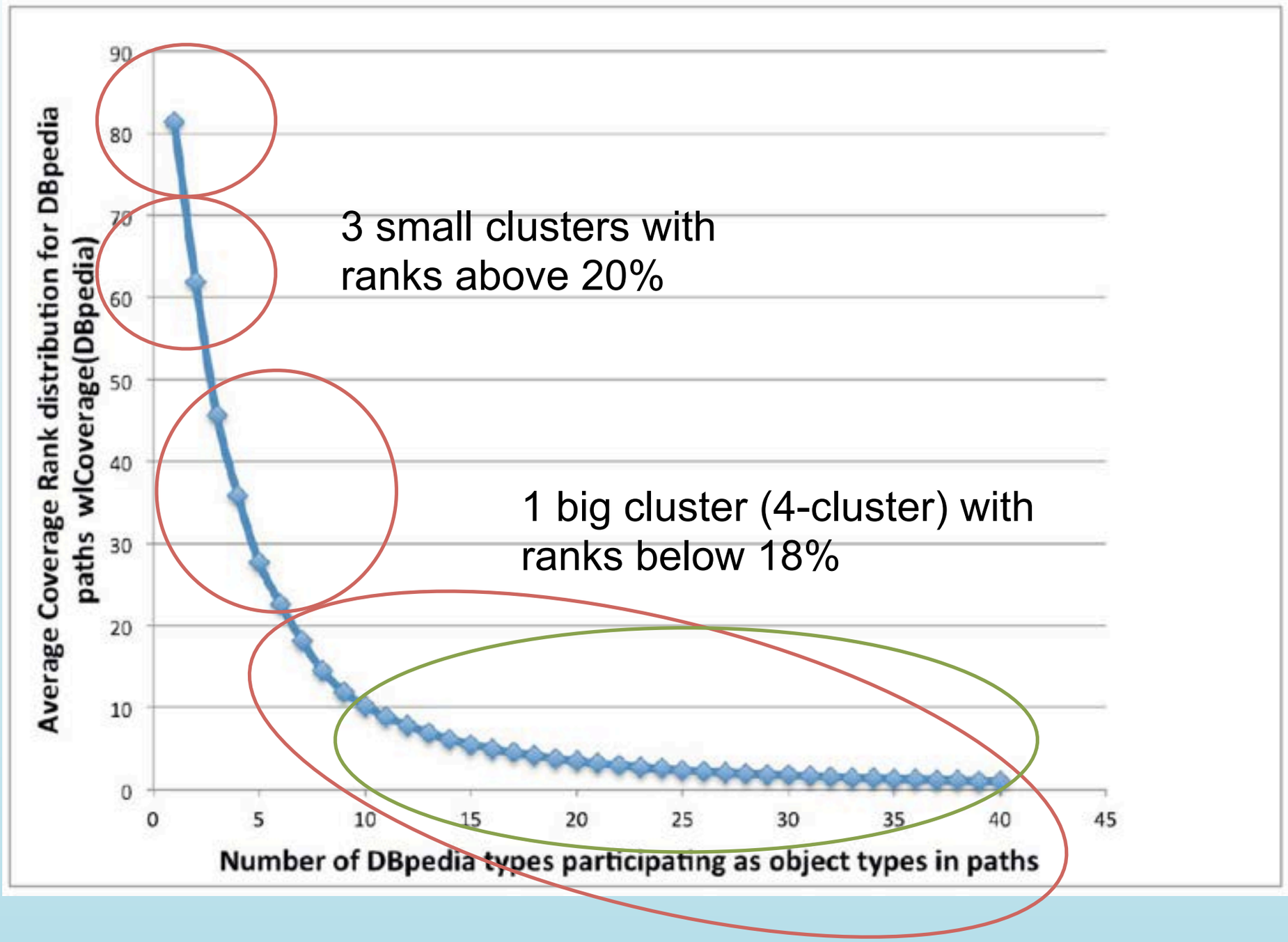


Encyclopedic Knowledge Patterns from Wikipedia Wikilinks (@ISWC2011)

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini: Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Ex.: *k*-means clustering on Wikipedia link popularity

Sample distribution of pathPopularity for DBpedia link paths. The y-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity

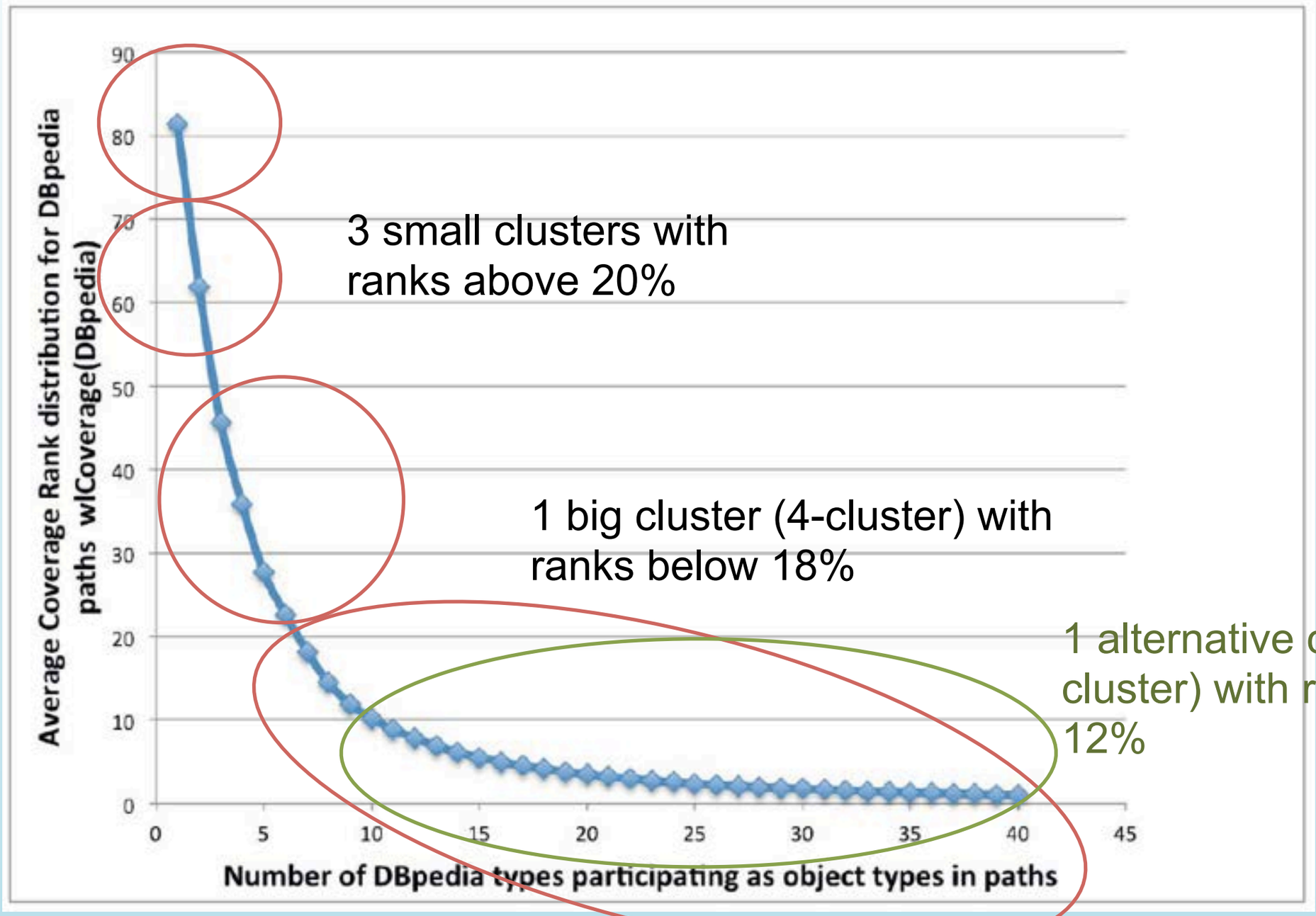


Encyclopedic Knowledge Patterns from Wikipedia Wikilinks (@ISWC2011)

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini: Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Ex.: *k*-means clustering on Wikipedia link popularity

Sample distribution of pathPopularity for DBpedia link paths. The y-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity

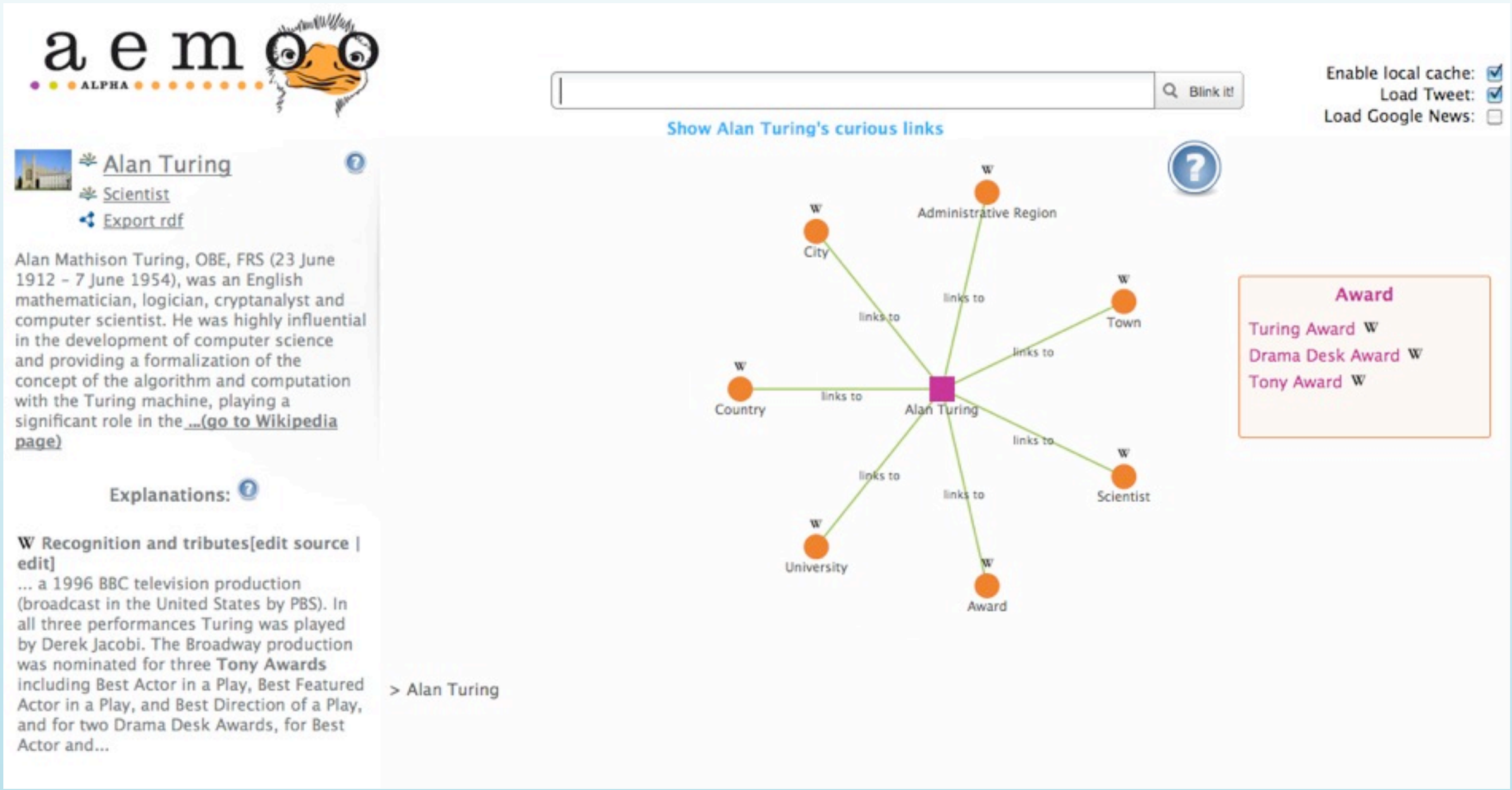


Encyclopedic Knowledge Patterns from Wikipedia Wikilinks (@ISWC2011)

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini: Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Serendipity in exploratory browsing

http://www.aemoo.org

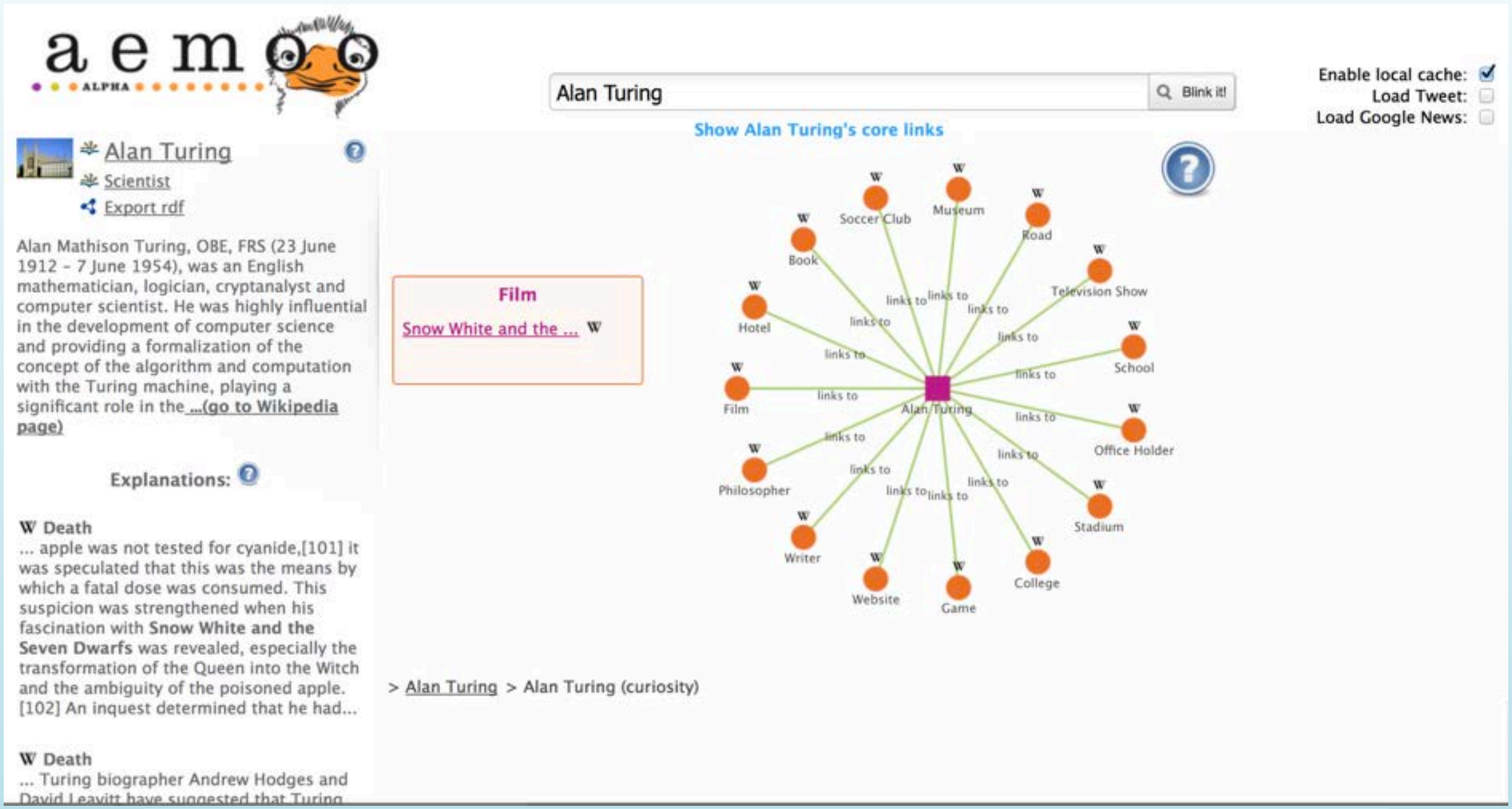


Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Alberto Musetti, Paolo Ciancarini: Aemoo: exploring knowledge on the web. *WebSci 2013*: 272-275

Aemoo: exploratory search based on EKP - *Semantic Web Challenge @ISWC 2011* – Short listed, 4th place

Serendipity in exploratory browsing

http://www.aemoo.org



Andrea Giovanni Nuzzolese, Valentina Presutti, Aldo Gangemi, Alberto Musetti, Paolo Ciancarini: Aemoo: exploring knowledge on the web. *WebSci 2013*: 272-275

Aemoo: exploratory search based on EKP - *Semantic Web Challenge @ISWC 2011*

invariances

semantic patterns

ontology design patterns

knowledge patterns

best practices

frames

semantic unit tests

requirements

interaction patterns

ontology design as problem solving

DL varieties

reasoning pipelines

query patterns

A broader vision: knowledge patterns and their façades

- The Knowledge Pattern (KP) Model is a discovery and analogy structure <*C,L,I,D,W,T,V,X,F,S,R,U*> such that a KP emerges out of invariances across multiple façades:
  - ***C***: concept graphs
  - ***L*:** logical forms (some axiomatization of *multigrade predicates*)
  - ***I*:** local inference rules (either classical or approximate)
  - ***D***: (relational) data
  - ***W***: linguistic data
  - ***T*:** social tagging data
  - ***V*:** interaction/visualization/formatting structures
  - ***X***: provenance data
  - ***F*:** framing information
  - ***S*:** sentic information
  - ***R*:** relations to other KP
  - ***U***: use cases

Informal graphs

DL & Rule (reasoning) patterns

Data patterns

Cognitive patterns

Ontology design patterns

Lexical and NLP data patterns

Socio-patterns

Web and interaction patterns

Task patterns

All façades can provide features for stochastic processes

All façades may be encoded in RDF for interoperability and joint reasoning

I'm in good company

- Peter Clark on Knowledge Patterns in 2001 (KR paper)
- Steffen Staab, Alan Rector, Vojtech Svatek, Chris Welty, Giancarlo Guizzardi, myself, Valentina Presutti, Eva Blomqvist, and many others proposing "semantic patterns", "ontology design patterns", etc. as means to pair ontology requirements and solutions
- Myself (ESWC2009 keynote): knowledge patterns as objects of empirical investigation
- Frank Van Harmelen (ISWC2011 keynote): route to empirical research: data science, data patterns
- Martin Hepp (EKAW2012 keynote): web semantics not necessarily coincident with DL and traditional OE, need for Web-oriented practices and patterns
- David Karger (ESWC2013 keynote): what can the SW do for average users? Not much until now, need for user-oriented patterns
- Enrico Motta (ESWC2013 keynote): what semantics in the current SW? Different forces, empirical KR to address human representing and reasoning with useful patterns
- John Sowa (SemTech2013 lecture): patterns exist at different levels of data, ontologies, and reality

... on the shoulders of

- Köhler, Bartlett, Piaget, Fillmore, Minsky, and many cognitive and neuro- scientists ...

Semantic web expressivity?

- Is our semantics enough to support extraction, representation, and harnessing of social semantics?
    - triples are simple structures
    - classes represent arbitrary concepts

where is cognitive adequacy?

when does a class represent arbitrary data, and when is it a counterpart of a human knowledge pattern?

is that difference important in general?

dbpedia-owl:areaTotal	127560000.000000 (xsd:double)
dbpedia-owl:country	dbpedia:Austria
dbpedia-owl:elevation	353 (xsd:double)
dbpedia-owl:postalCode	A-801x, A-802x, A-803x, A-804x, A-805x
dbpprop:website	http://www.graz.at
dbpprop:elevation	353 (xsd:integer)
dbpprop:mayor	Siegfried Nagl
dbpprop:state	dbpedia:Styria
dbpprop:district	dbpedia:Statutory_city

DBpedia

When triplifying Wikipedia infoboxes, its designers lost the framing of boxes and internal sub-boxes

The case of Infoboxes

- Infobox framing is missing in the DBpedia ontology too
- If we mine the ontology to check what properties can be applied to what classes, the result is partial and often non-correspondent to the original frame
- Scraping heuristics may be more cognitively-sound ...

Interaction semantics

- Interfaces and interaction patterns convey frame semantics
  - Schema induction and triplification of databases can be improved by exploiting interfaces exposing data, cf. data.cnr.it ontology design
  - HTML pages and stylesheets contain a lot of framing knowledge, cf. Craig Knoblock's work
  - Infographics can change the way we interpret the same data

Empirical conservativeness

What is present with a function in (evident, extracted, emerging) empirical data should be preserved in its semantic representation

Empirical conservativeness

- It is a measure against "oversimplification"
- The case of Infobox framing loss is a sample violation of this principle

Special case

- Keeping *interaction boundaries* is a special case of empirical conservativeness
- Like neural binding (at the neural level) and linguistic framing (at the cognitive level), relevant boundaries of logical representations need to be represented

Cf. original Marvin Minsky's frames:

"representations that mirror cognitive mechanisms"

More semantics or more distinctions?

- I am not advocating for "more semantics" in terms of complexity, rather for more distinctions
  - Human knowledge is relational in nature
  - We need n-ary and multigrade relations, but arbitrary relations would be too much in current KR scenarios, then we can use them with *smart reification patterns*
  - Classes are powerful primitives in logical languages, specially in description logics and triple-based languages

More semantics or more distinctions?

- Fixation on classes goes with a trade-off
  - Classes need to be distinguished in terms of design
  - Class-oriented representation needs a “push-up” to partly recover the lost structure

Class types?

Types of classes have been distinguished in the past

- AI: sorts and types
- Formal Ontology: OntoClean metaclasses, based on formal criteria
- OWL2 punning: arbitrary typing of classes

Solutions span between the two extremes: heavy principles (OntoClean) - no principle at all (punning)

Relevance in modelling

- "What's special in that class?"
  - E.g., *it's central in the data, it's a frame, it's an n-ary reification mechanism, it's the result of a discovery algorithm, etc.*
- A new vocabulary for metaclasses?

Is there anything like that in OWL or RDF?

Maybe ontology modules, `classes`, `named graphs`, `hasKey` axioms

Not specific to the boundary problem, nor to framing or neural binding

**Very recent: new spec for named graphs accepts typing*

- Linked Open Vocabularies is a good starting point to find out elements of vocabularies that are useful
- E.g. we are interested in “events”
  - http://lov.okfn.org/dataset/lov/search/#s=event

Linked Open Vocabularies (LOV)

developped by Pierre-Yves Vandenbussche

The "LOV Search" Features gives you the possibility to search for an existing element (property, class or vocabulary) in the Linked Open Vocabularies Catalogue.
LOV Aggregator endpoint and metrics about the use of vocabularies in the Semantic Web are used to bring you some relevant results.

Endpoint

event Search

Filter by Domain

- City (143)
- Data & Systems (26)
- General (168)
- Library (78)
- Market (29)
- Media (146)
- Metadata (32)
- Science (68)

Filter by Type

- rdfs:Class (412)
- rdf:Property (472)
- voaf:Vocabulary (23)
- Other (90)

Filter by Vocabulary (96)

- bio (64)
- schema (64)
- crm (42)
- oc (36)
- swc (33)
- ncal (33)
- mo (27)
- next (25)

863 results in 96 vocabularies

event (voaf:Vocabulary) score:0.826
rdfs:label The **Event** ontology
dcterms:title The **Event** Ontology @en
dce:title The **Event** ontology
dcterms:description ...e notion of reified **events** - events seen as f... @en
dce:description ...e notion of reified **events**--- events seen a.....ines one concept: **Event**, which may have age.....tological status of **event** tokens, they are se...
bibo:shortTitle **Event** @en
vann:preferredNamespacePrefix **event**
vann:preferredNamespaceUri .../purl.org/NET/c4dm/**event**.owl#

event:Event (owl:Class) score:0.607
rdfs:label **Event**
rdfs:label **Event** @en
rdfs:comment ...cognitive agent. An **event** may have actively p... @en
rdfs:comment ...cognitive agent. An **event** may have actively p...

crm:E5_Event (rdfs:Class) score:0.606
rdfs:label **Evento** @pt
rdfs:label **Event** @en
rdfs:comment ...ction between an E5 **Event** and an E4 Period is.....el of detail, an E5 **Event** is an 'instantaneou..... fine level, the E5 **Event** can be analysed int... @en

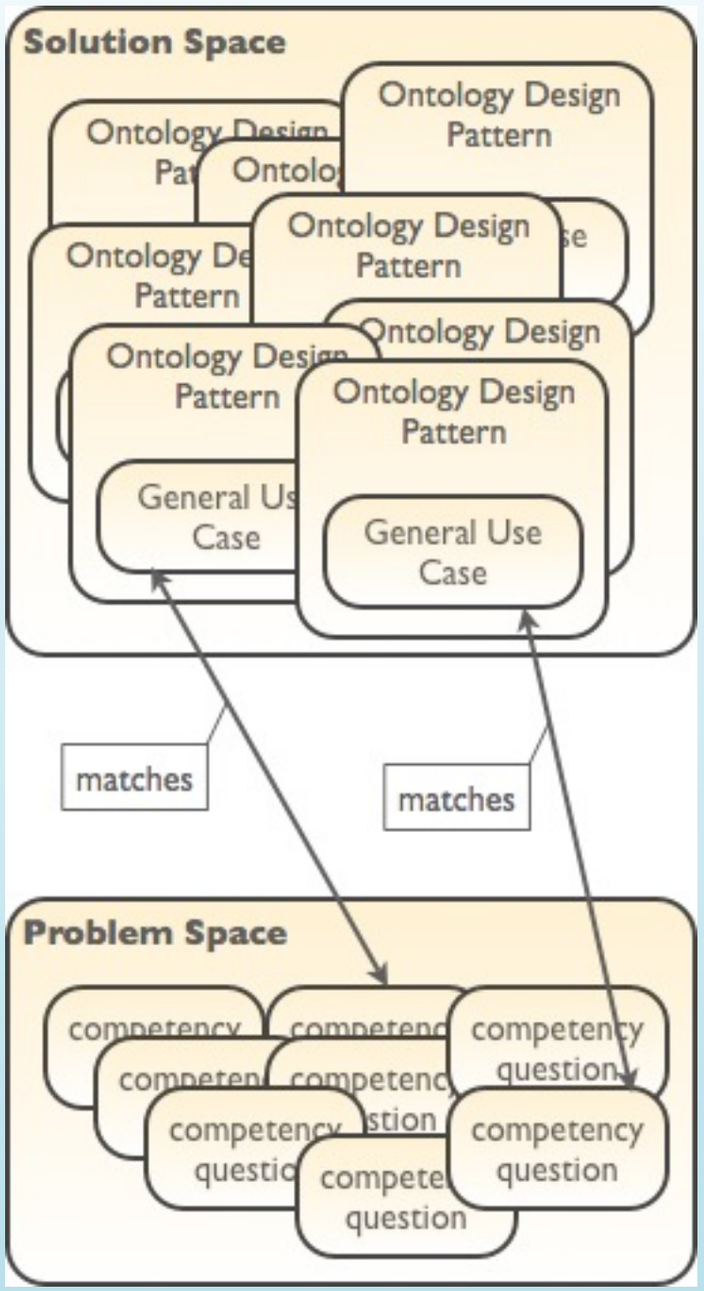
bio:Event (owl:Class) score:0.606
rdfs:label **Evento** @es
rdfs:label **Event** @en
rdfs:comment An **event** is an occurrence th.....nd/or other agents. **Events** are assumed to occ... @en
vann:usageNote ...scribe biographical **events**, i.e. events in th... @en

dul:Event (owl:Class) score:0.606
rdfs:label **Evento** @it
rdfs:label **Event** @en
rdfs:comment ... or mental process, **event**, or state. More th...eoretically, **events** can be classified .....owing explains why: **events** are related to obs.....d classification of **events** seems the most sta.....ws Consider a same **event** 'rock erosion in th.....ent), as a punctual **event** (if we collapse the.....s refer to the same **event**, but are still diff.....ion) is to classify **events** based on aspectual.....entities for a same **event**, where the

- Overcoming OE bottlenecks, how good design and automatic extraction can have a date together
- Useful KP-based abstraction from text, data, or ontologies
- KP directly reusable at design time

Ontology Design Patterns

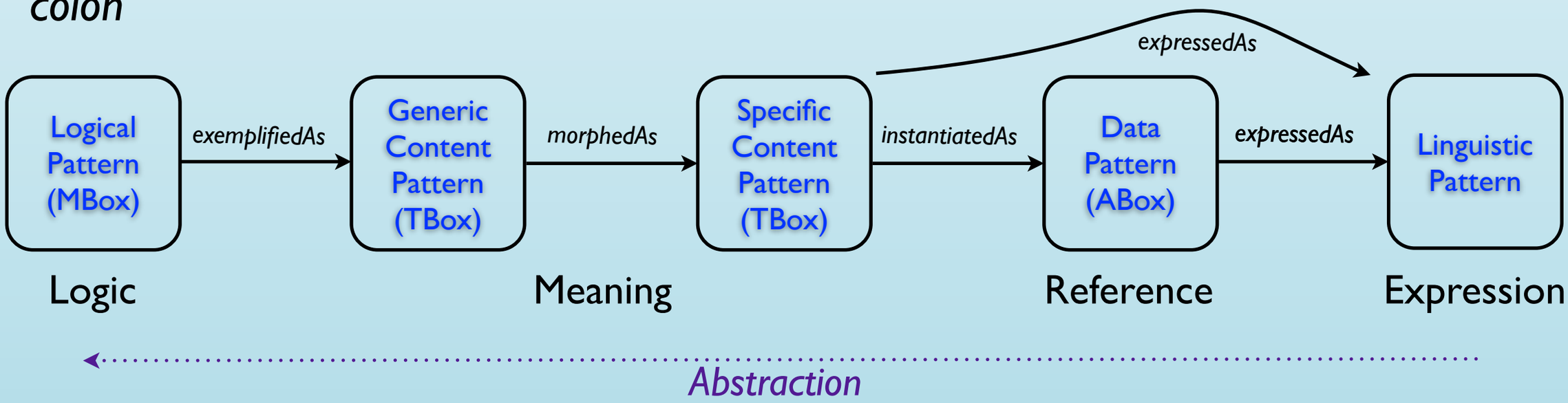
An ontology design pattern is a reusable successful solution to a recurrent modeling problem



Layered pattern morphisms

An ontology design pattern describes a formal expression that can be *exemplified*, *morphed*, *instantiated*, and *expressed* in order to solve a domain modelling problem

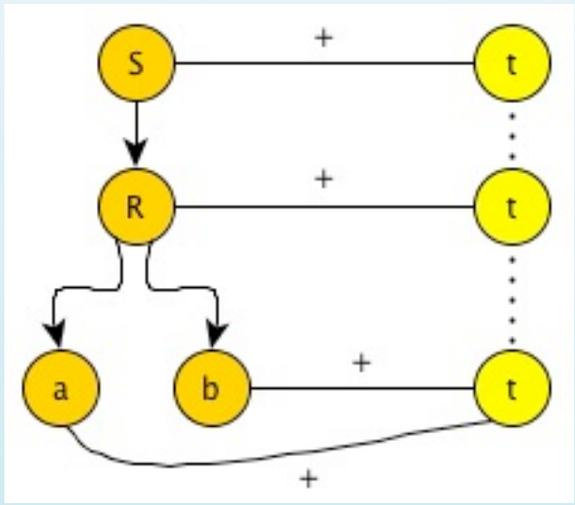
- **owl:Class:_:x rdfs:subClassOf owl:Restriction:_:y**
- *Inflammation rdfs:subClassOf (localizedIn some BodyPart)*
- *Colitis rdfs:subClassOf (localizedIn some Colon)*
- *John's_colitis isLocalizedIn John's_colon*
- *"John's colon is inflammated", "John has got colitis", "Colitis is the inflammation of colon"*



Aldo Gangemi, Valentina Presutti: Ontology Design Patterns. *Handbook on Ontologies 2nd ed.* (2009)

Design going empirical: N-ary patterns in KR

- Temporal indexing pattern
  - (R(a,b))+t *sentence indexing*
    - quads, external time stamps
  - R(a,b)+t *relation indexing*
    - reified n-ary relations (3D frames)
  - R(a+t,b+t) *individual indexing*
    - fluents, 4D, tropes, "context slices" (4D frames)
  - tR *name nesting*
    - ad hoc naming of binary relations



- More indexes for additional arguments

Aldo Gangemi, Valentina Presutti: A Multi-dimensional Comparison of Ontology Design Patterns for Representing n-ary Relations. SOFSEM 2013: 86-105
Andreas Scheuermann, Enrico Motta, Paul Mulholland, Aldo Gangemi and Valentina Presutti. An Empirical Perspective on Representing Time. K-CAP 2013

Abstraction going empirical: centrality discovery in datasets

mo:Track

mo:Playlist

mo:track

mo:Record

mo:available_as

mo:MusicArtist

foaf:maker

mo:ED2K

mo:available_as

tags:taggedWithTag

mo:available_as

mo:Torrent

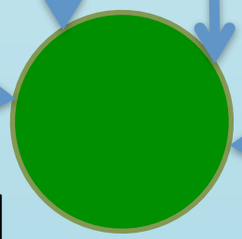
tags:Tag

mo:image

dc:date

dc:title

dc:description

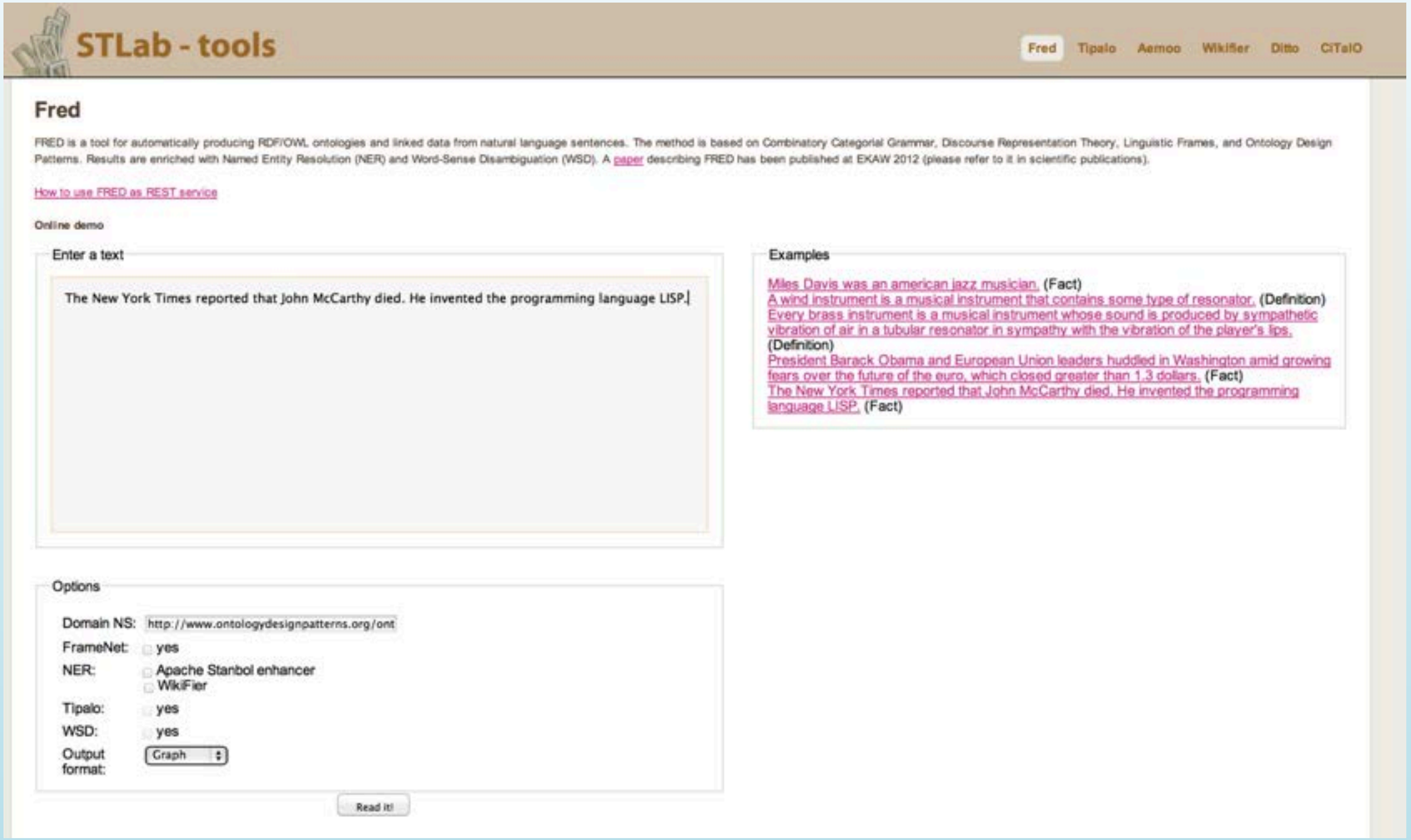


rdfs:Literal

Valentina Presutti, Lora Aroyo, Alessandro Adamou, Balthasar Schopman, Aldo Gangemi, Guus Schreiber: Extracting Core Knowledge from Linked Data. COLD2011, CEUR-WS.org Vol-782.

Abstraction going empirical: machine reading with FRED

http://wit.istc.cnr.it/stlab-tools/fred/

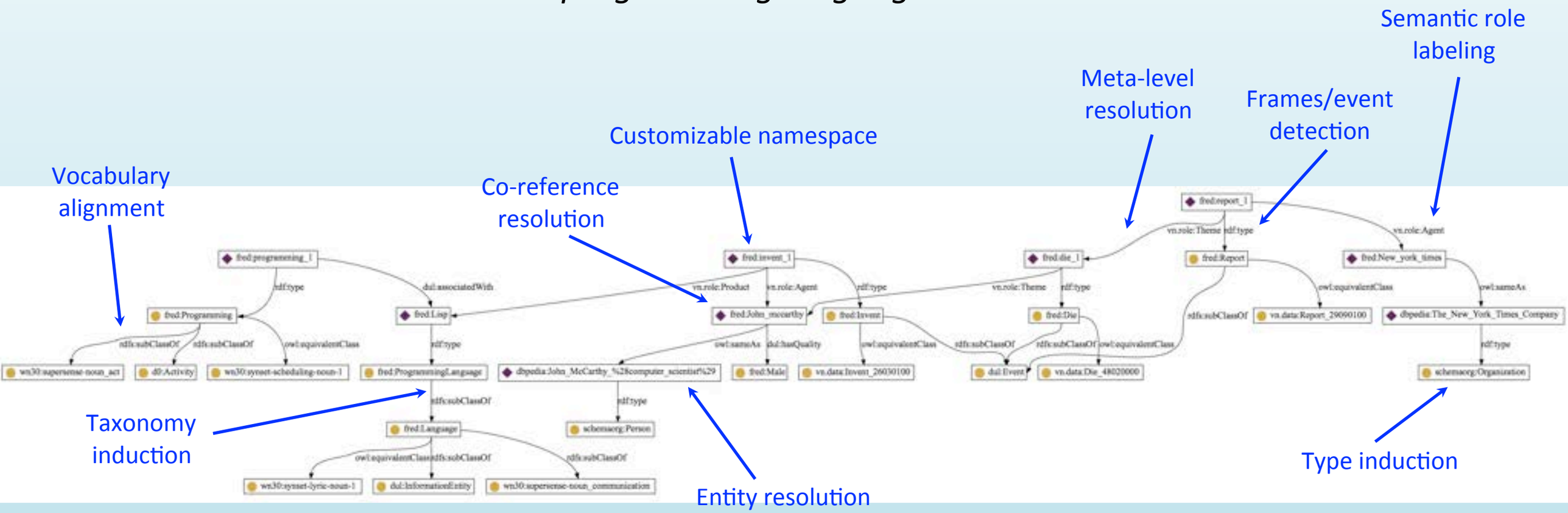


Valentina Presutti, Francesco Draicchio, Aldo Gangemi: Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames. *EKAW 2012*: 114-129

Event and Frames from text

http://wit.istc.cnr.it/stlab-tools/fred/

The New York Times reported that John McCarthy died. He invented the programming language LISP.



Valentina Presutti, Francesco Draicchio, Aldo Gangemi: Knowledge Extraction Based on Discourse Representation Theory and Linguistic Frames. *EKAW 2012*: 114-129