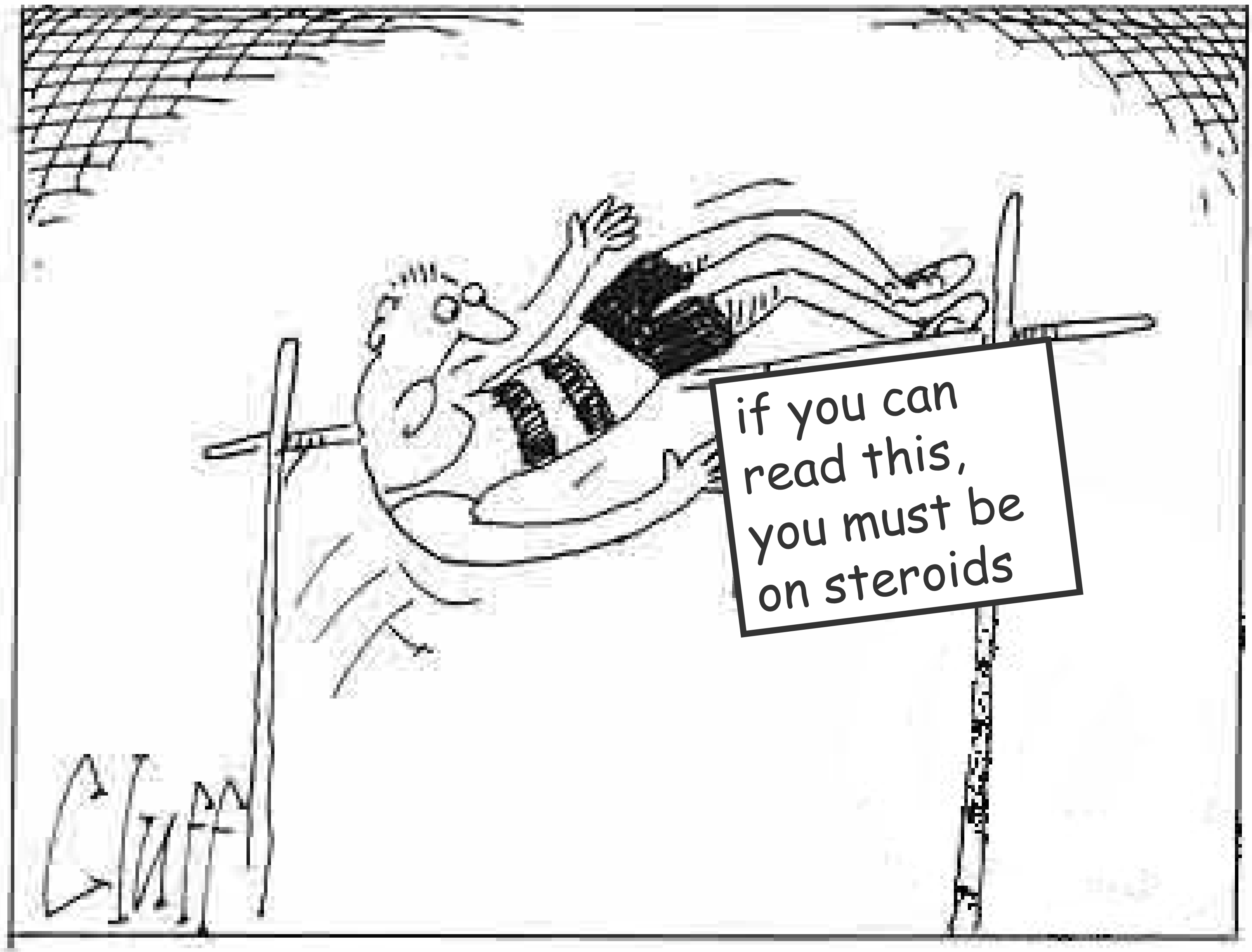
Where does it break? or:

Why Semantic Web research is not just "Computer Science as usual"

Frank van Harmelen
AI Department
Vrije Universiteit Amsterdam

vrije Universiteit


if you can
read this,
you must be
on steroids
Cliff

Oh no, not more “vision”...

Don’t worry,
there will be lots of technical content

Grand Topics…

- what are the **science challenges** in SW?
- Which **implicit traditional assumptions** break?
- Illustrated with 4 such “traditional assumptions”

and also:

- “Which Semantic Web” ?

Before we go on:

Which Semantic Web are we talking about?

Which Semantic Web?

- Version 1:
  "Semantic Web as Web of Data" (TBL)

- **recipe**:
  expose databases on the web,
  use RDF, integrate
- **meta-data** from:
  - expressing DB schema semantics
    in machine interpretable ways
- **enable** integration and unexpected re-use

Which Semantic Web?

- Version 2:
  "Enrichment of the current Web"

- **recipe**:
  Annotate, classify, index
- **meta-data from**:
  - automatically producing markup:
    named-entity recognition,
    concept extraction, tagging, etc.
- **enable** personalisation, search, browse,..

Which Semantic Web?

- Version 1:
  "Semantic Web as Web of Data"

- Version 2:
  "Enrichment of the current Web"

- Different use-cases
- Different techniques
- Different users

Semantic Web:

Science or technology?

Semantic Web as Technology

- better search & browse
- personalisation
- semantic linking
- semantic web services
- ...

Semantic Web as Science

4 examples of "where does it break?"

- old assumptions that no longer hold,
- old approaches that no longer work

4 examples of “where does it break?”

❶ Traditional complexity measures

Who cares about decidability?

- Decidability ≈ **completeness**
  guarantee to find an answer,
  or tell you it doesn't exist,
  given enough run-time & memory
- **Sources of incompleteness**:
  - incompleteness of the input data
  - insufficient run-time to wait for the answer

➔ **Completeness is unachievable in practice anyway**,
regardless of the completeness of the algorithm

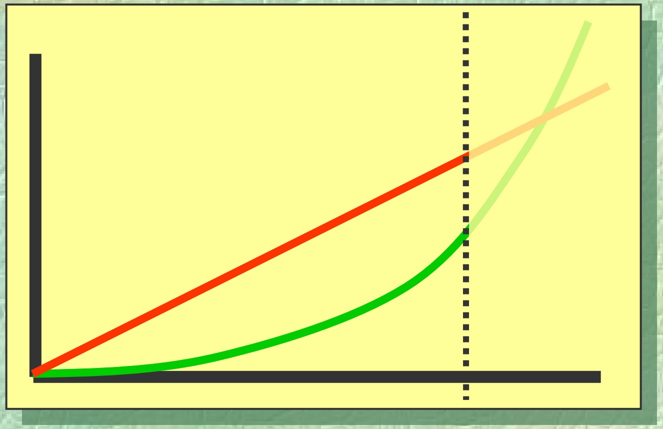
Who cares about undecidability?

- Undecidability
  ≠ always guaranteed not to find an answer
- Undecidability
  = not always guaranteed to find an answer

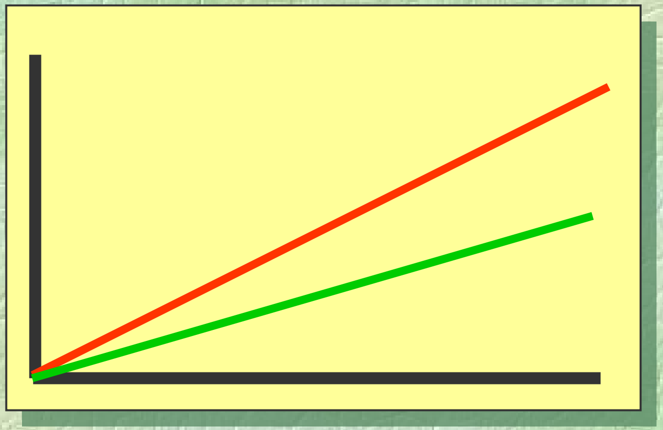
- **Undecidability may be harmless**
  in many cases;
  in all cases that matter

Who cares about complexity?

- **worst-case**: may be exponentially rare
- **asymptotic**

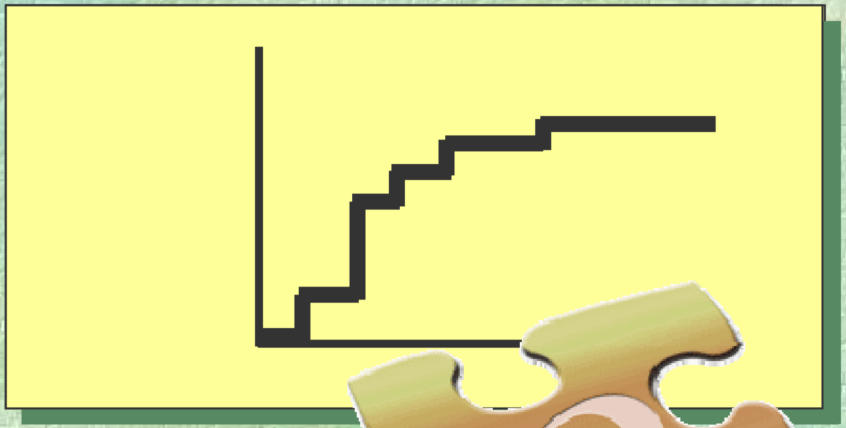


- **ignores constants**



What to do instead?

- Practical observations on RDF Schema:
  - Compute full closure of $O(10^9)$ statements
- Practical observations on OWL:
  - NEXPTIME ☹
    but fine on many practical cases☺
- Do more experimental performance profiles with realistic data



- Think hard about **"average case" complexity….**

4 examples of "where does it break?"

❶ Traditional complexity measures

❷ Hard in theory, easy in practice

Example: Reasoning with Inconsistent Knowledge

This work with
Zhisheng Huang &
Annette ten Teije

Knowledge will be inconsistent

Because of:

- mistreatment of defaults
- homonyms
- migration from another formalism
- integration of multiple sources

New formal notions are needed

- New notions:
  - **Accepted**: $T \mathrel{|\!\approx} \phi$ and $T \mathrel{|\!\not\approx} \neg\phi$
  - **Rejected**: $T \mathrel{|\!\not\approx} \phi$ and $T \mathrel{|\!\approx} \neg\phi$
  - **Overdetermined**: $T \mathrel{|\!\approx} \phi$ and $T \mathrel{|\!\approx} \neg\phi$
  - **Undetermined**: $T \mathrel{|\!\not\approx} \phi$ and $T \mathrel{|\!\not\approx} \neg\phi$

- **Soundness**: (only classically justified results)

$$T \mathrel{|\!\approx} \phi \Rightarrow (\exists T' \subseteq T)(T' \not\models \perp \text{ and } T' \models \phi)$$

Basic Idea

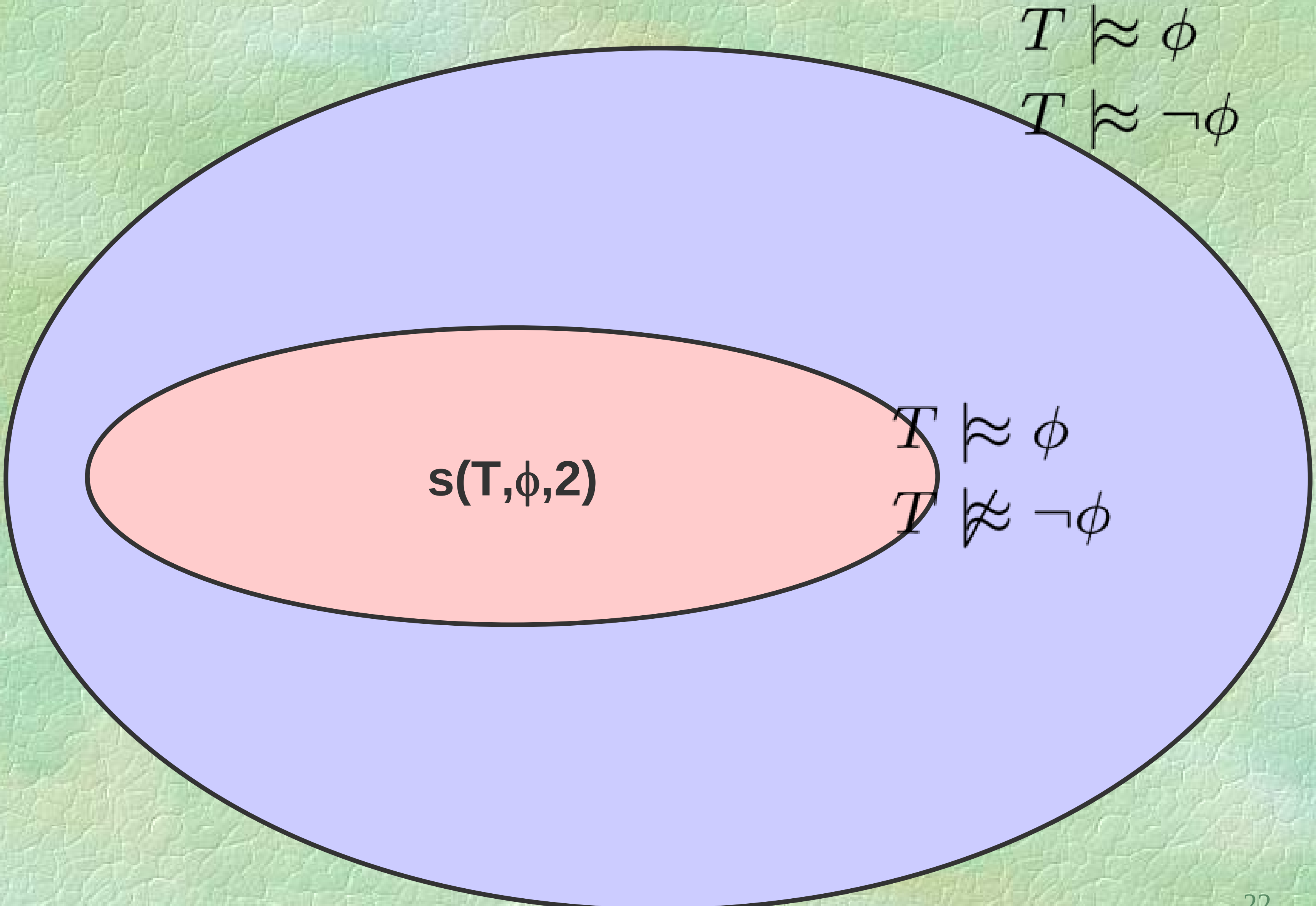
1. Start from the query
2. Incrementally select larger parts of the ontology that are "relevant" to the query, until:
   i. you have an ontology subpart that is small enough to be consistent and large enough to answer the query

   or

   ii. the selected subpart is already inconsistent before it can answer the query

Selection function

General Framework



More precisely:

Use selection function $s(T,\phi,k)$,
with $s(T,\phi,k) \subseteq s(T,\phi,k+1)$

1. Start with k=0:
   $s(T,\phi,0) \mid\approx \phi$ or $s(T,\phi,0) \mid\approx \neg\phi$?
2. Increase k, until
   $s(T,\phi,k) \mid\approx \phi$ or $s(T,\phi,k) \mid\approx \neg\phi$
3. Abort when
   - undetermined at maximal k
   - overdetermined at some k

Nice general framework, but...

- which selection function $s(T,\phi,k)$ to use?
- Simple option: **syntactic distance**
  - put all formulae in clausal form:
    $a_1 \vee a_2 \vee \ldots \vee a_n$
  - **distance k=1** if some clausal letters overlap
    $a_1 \vee X \vee \ldots \vee a_n$, $b_1 \vee \ldots X \vee b_n$
  - **distance k** if chain of k overlapping clauses are needed
    $a_1 \vee X \vee \ldots X_1 \vee a_n$
    $b_1 \vee X_1 \vee \ldots X_2 \vee b_n$,
    ….
    $c_1 \vee X_k \vee \ldots X \vee c_n$

Evaluation

Ontologies:

- Transport: 450 concepts
  Communication: 200 concepts
  Madcow: 55 concepts

Selection functions:

- symbol-relevance = axioms overlap by ≥1 symbol
- concept-relevance ≈ axioms overlap by ≥1 concept

Query a random set of subsumption queries:

$$\text{Concept1} \subseteq \text{Concept2} \text{ ?}$$

Evaluation: Lessons

ontology	relevance	queries	Intend.	Caut.	Reckl.	counter intuit.	Intend. %
MadCow+	symbol	2594	2538	0	54	2	98%
	concept	2594	2402	192	0	0	93%
Commun.	symbol	6576	6396	8	164	8	97%
	concept	6576	6330	246	0	0	96%
Transport	symbol	6258	5504	0	752	2	88%
	concept	6258	6228	30	0	0	99%

this makes concept-relevance a high quality sound approximation (> 90% recall, 100% precision)

Works surprisingly well

On our benchmarks,
allmost all answers are "intuitive"

- Not well understood why
- Theory doesn't predict that this is easy
  - paraconsistent logic,
  - relevance logic
  - multi-valued logic
- Hypothesis:
  due to "local structure of knowledge"?

4 examples of "where does it break?"

❶ Traditional complexity measures

❷ Hard in theory, easy in practice

❸ context-specific nature of knowledge

Opinion poll

left

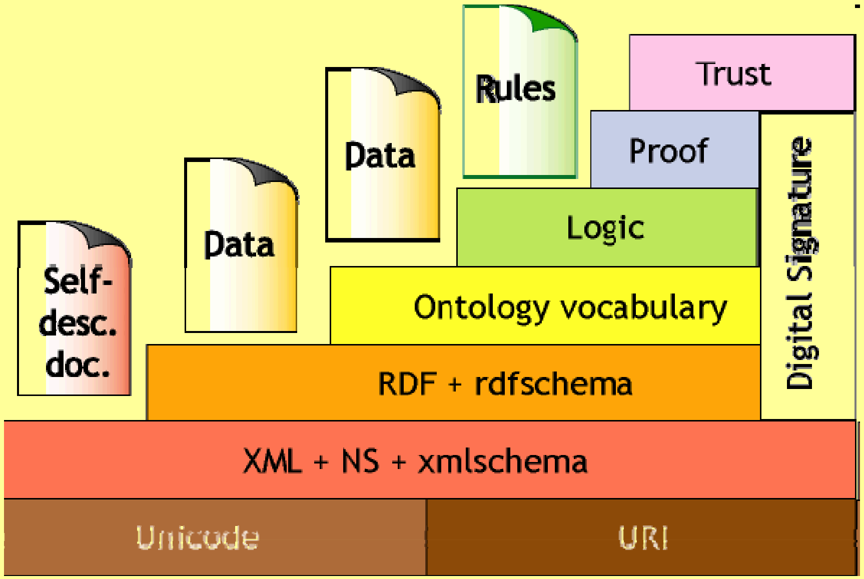
meaning of a sentence
is **only** determined
by the sentence itself,
and **not** influenced by
the surrounding
sentences,
and **not** by the situation
in which the sentence
is used

right

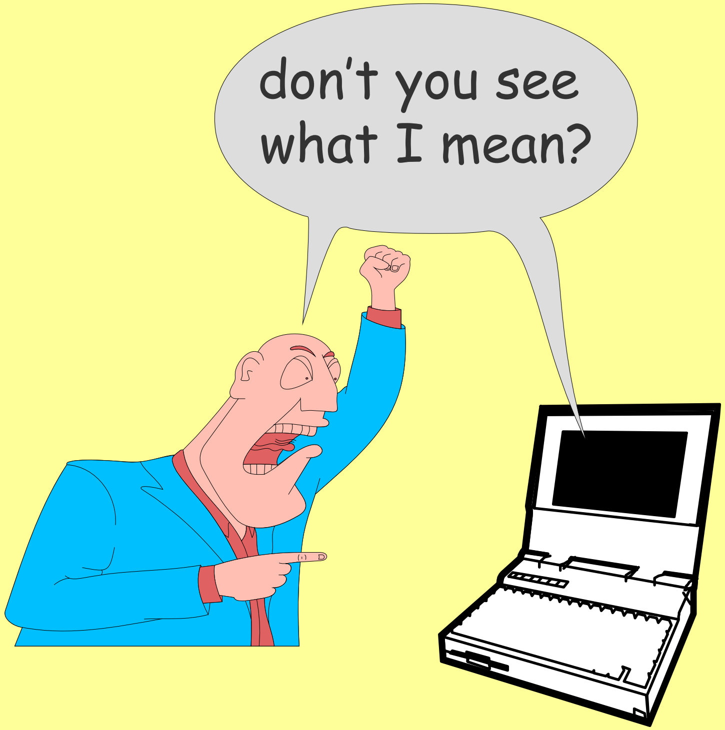
meaning of sentence
is **not only** determined
by the sentence itself,
but is **also** influenced by
by the surrounding
sentences,
and **also** by the situation
in which the sentence
is used

Opinion poll

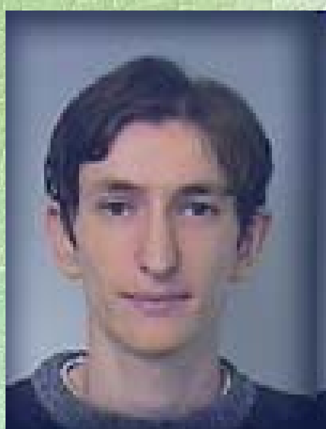
left



right



Example: Ontology mapping with community support

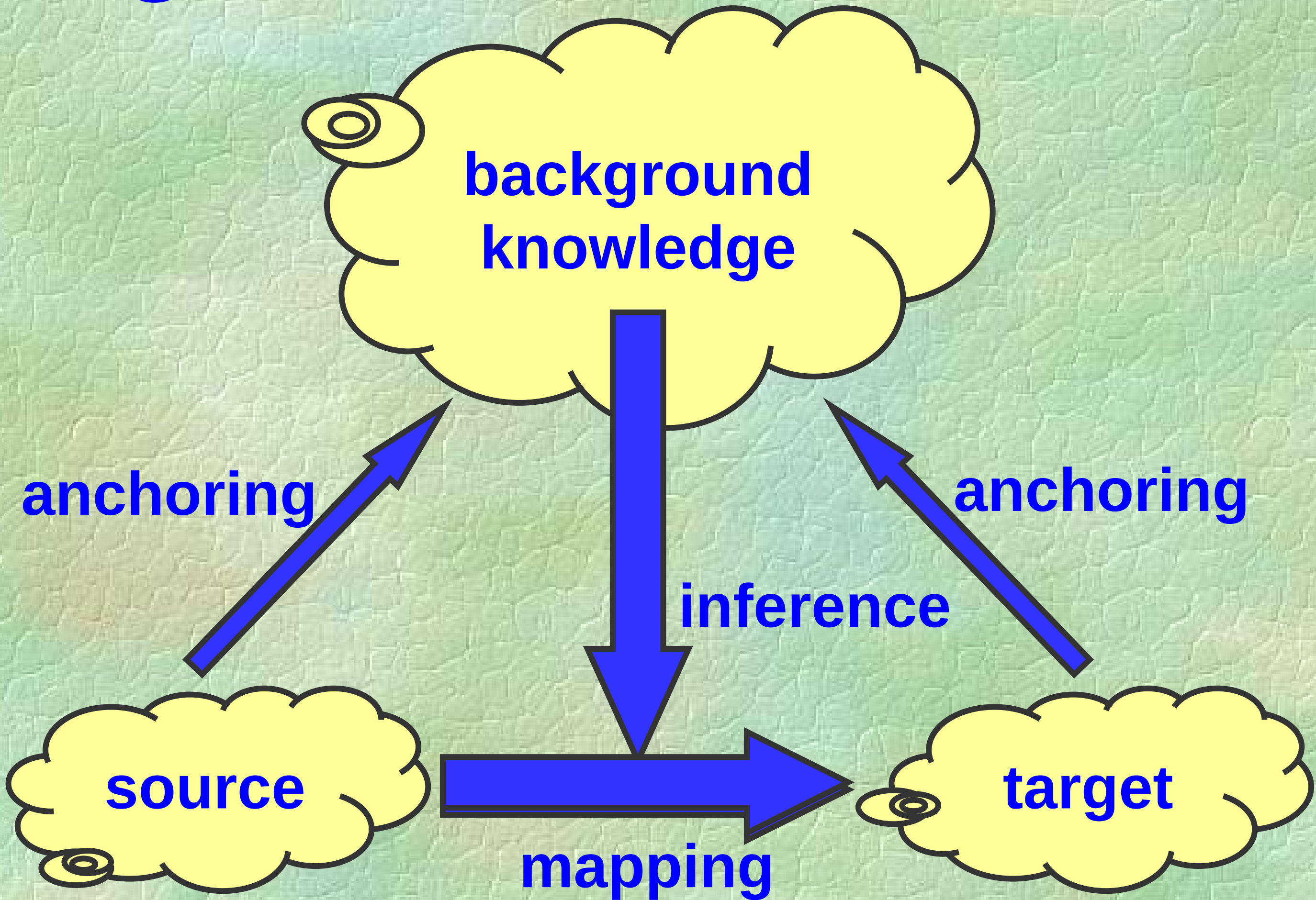


This work with

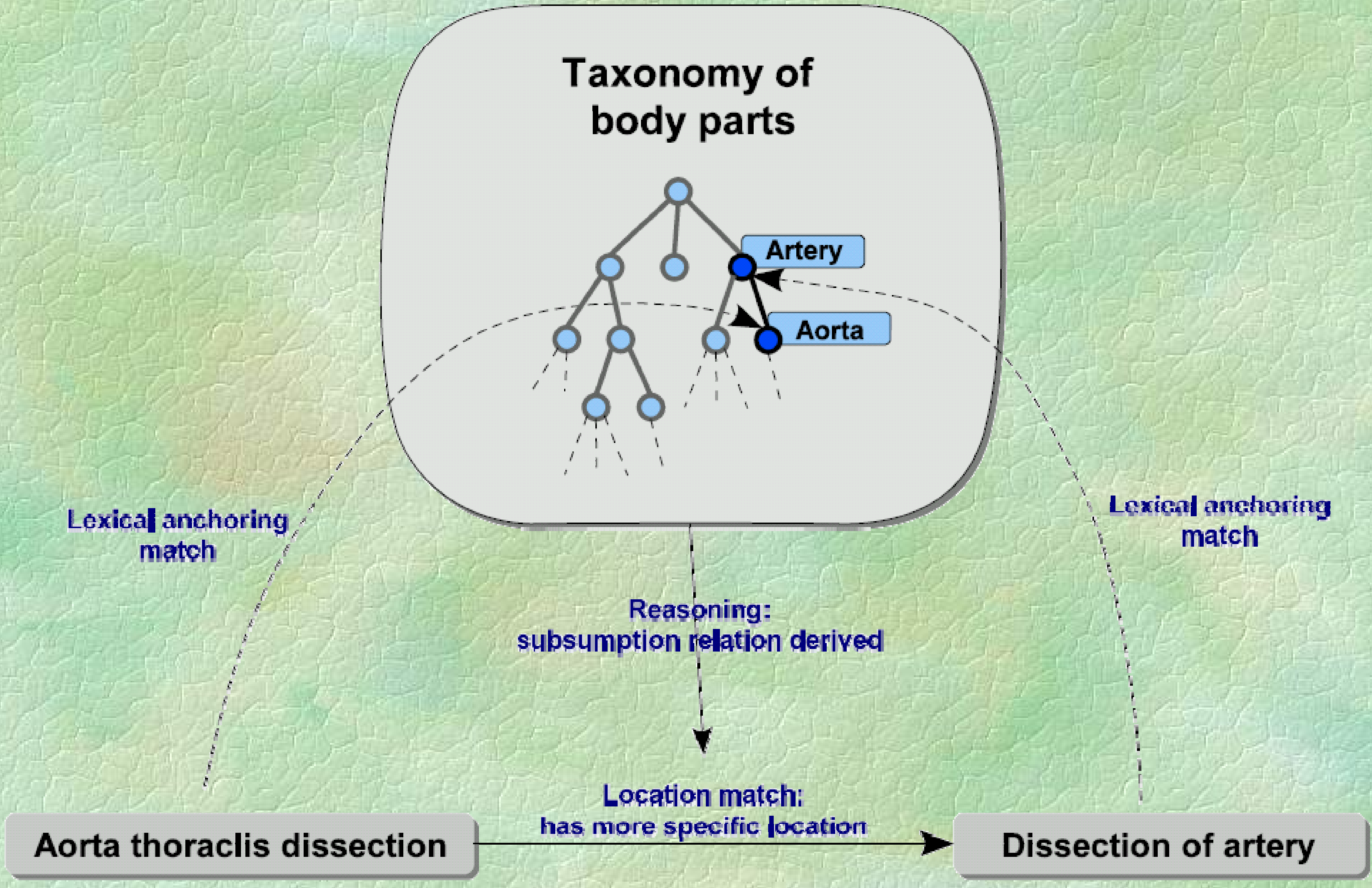
Zharko Aleksovski &

Michel Klein

The general idea

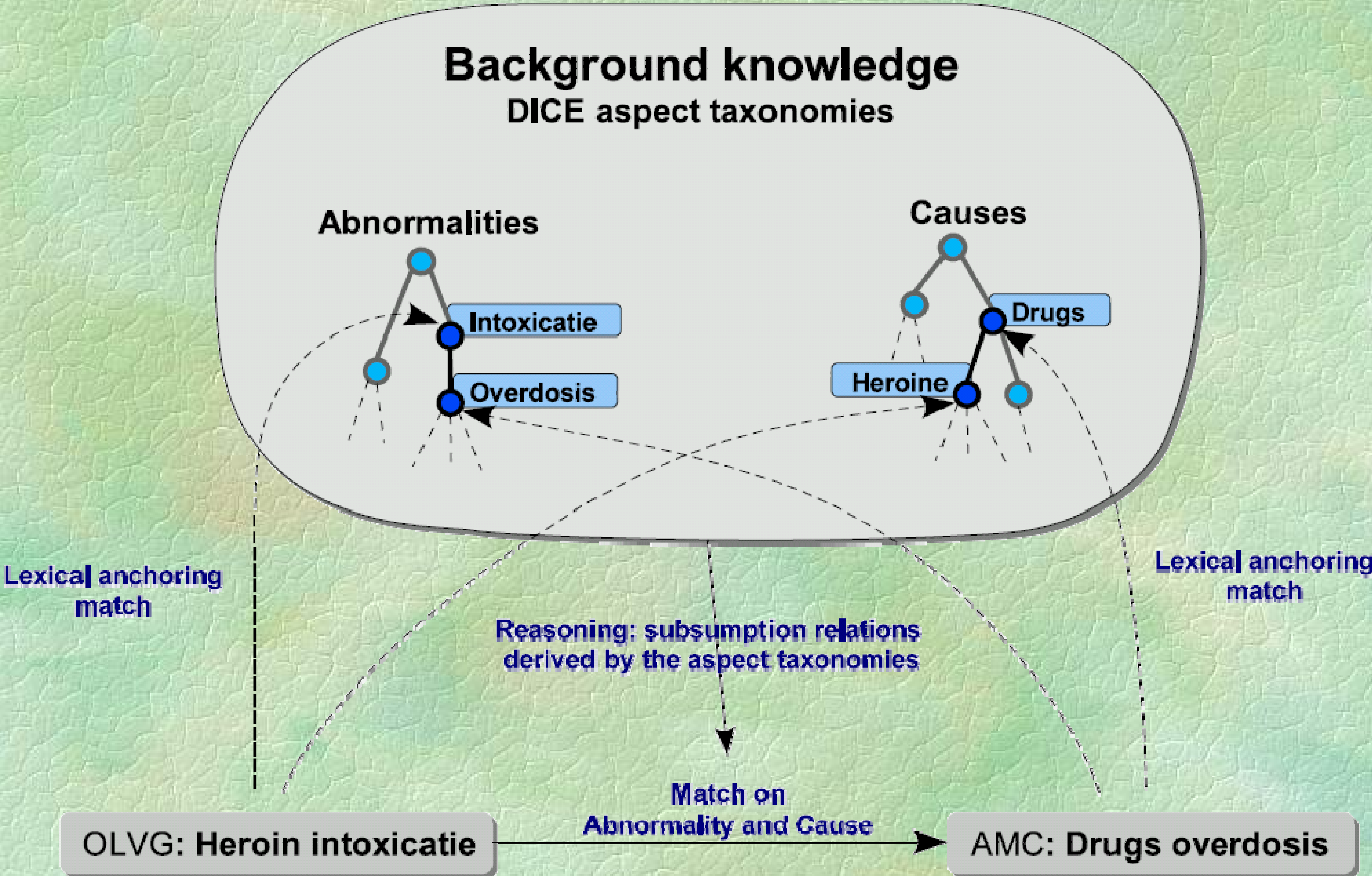


Example 1



Taxonomy of body parts
Artery
Aorta
Lexical anchoring match
Reasoning: subsumption relation derived
Lexical anchoring match
Location match: has more specific location
Aorta thoraclis dissection
Dissection of artery

Example 2


Background knowledge
DICE aspect taxonomies
Abnormalities
Intoxicatie
Overdosis
Causes
Drugs
Heroine
Lexical anchoring match
Lexical anchoring match
Reasoning: subsumption relations derived by the aspect taxonomies
Match on Abnormality and Cause
OLVG: Heroin intoxicatie
AMC: Drugs overdosis

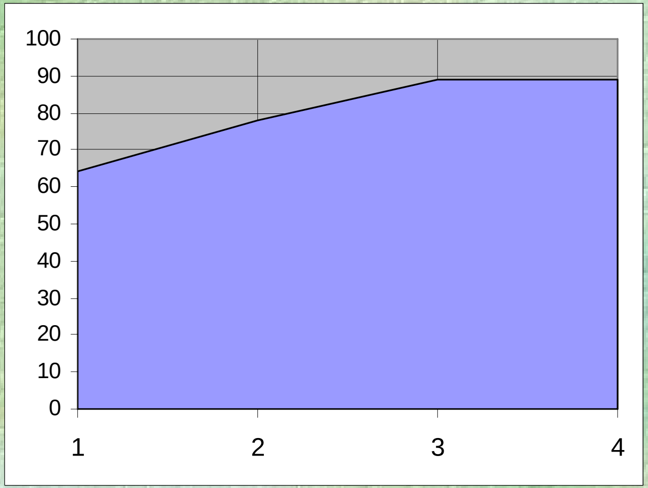
Experimental results

- Source & target = **flat lists** of ±1400 ICU terms each
- Anchoring = substring + simple germanic morphology
- Background = DICE (2300 concepts in DL)

	Semantic matching	Own Lexical matching	FOAM	Falcon-AO
agreement on single best match	65 (=32%)	43	35	22
agreement among top 5 matches	8 (= 4%)			
agreement on no match possible	43 (=22%)	43	26	32
improvement over expert match	35 (18%)	6	6	6
TOTAL POSITIVE:	**151 (=76%)**	**92 (=46%)**	**67 (=33%)**	**60 (=30%)**
wrong match found		5	47	78
incorrectly found no match	49(=24%)	103	86	62
TOTAL NEGATIVE:	**49(=24%)**	**108 (=54%)**	**133 (=67%)**	**140 (=70%)**

New results:

- more background knowledge makes mappings better
  - DICE (2300 concepts)
  - MeSH (22000 concepts)
  - ICD-10 (11000 concepts)
- Monotonic improvement of quality
- Linear increase of cost



So…

- The OLVG & AMC terms get their meaning from the context in which they are being used.
- Different background knowledge would have resulted in different mappings
- Their semantics is not context-free
- See also: S-MATCH by Trento

4 examples of
“where does it break?”

➊ Traditional complexity measures

➋ Hard in theory, easy in practice

➌ context-specific nature of knowledge

➍ logic vs. statistics

Logic vs. statistics

- DB schema's & integration is **only logic**, no statistics

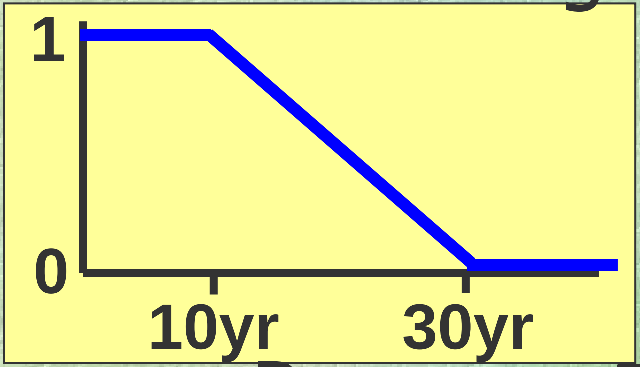
- AI is both logic and statistics, but **completely disjoint**

- Find **combinations** of the two worlds?
  - Statistics in the logic?
  - Statistics to control the logic?
  - Statistics to define the semantics of the logic?

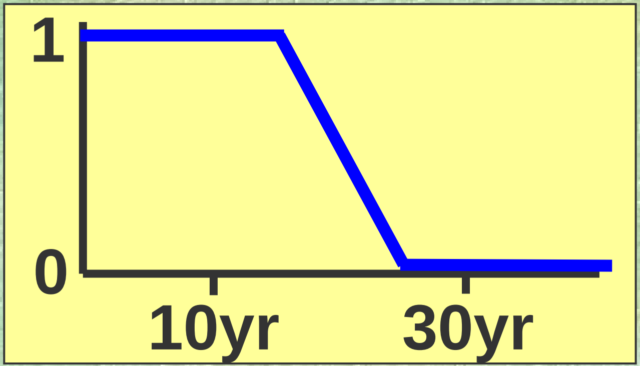
Statistics in the logic? Fuzzy DL

- (TalksByFrank $\sqsubseteq$ InterestingTalks) $\geq$ 0.7
- (Turkey:EuropeanCountry) $\leq$ 0.2
- youngPerson = Person $\sqcap$ $\exists$ age.Young

  Young(x) =



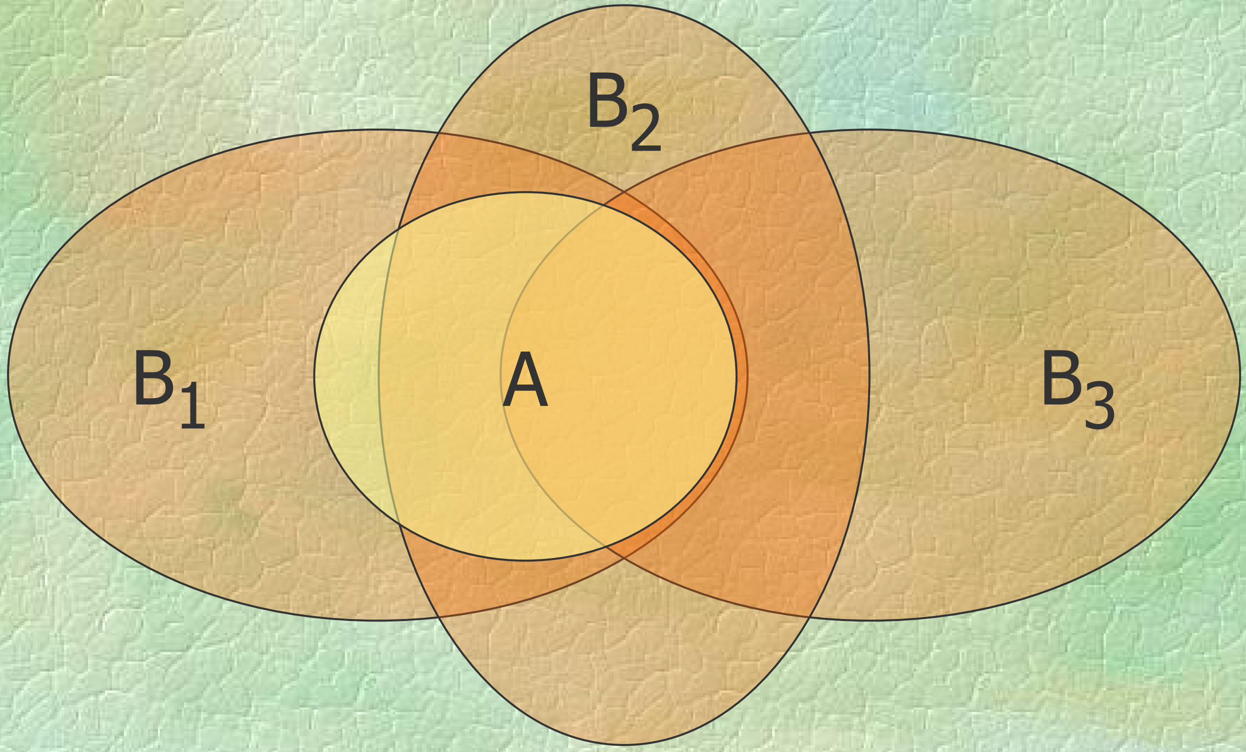
- veryYoungPerson = Person $\sqcap$ $\exists$ age.very(Young)



Umberto Straccia

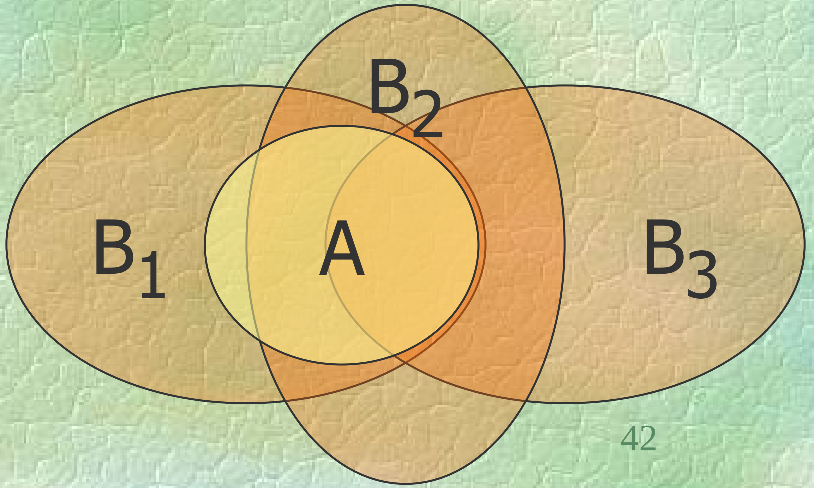
Statistics to control the logic?

- query: $A \sqsubseteq B$?
- $B = B_1 \sqcap B_2 \sqcap B_3 \rightarrow$ $\boxed{A \sqsubseteq B_1}$, $A \sqsubseteq B_2$, $A \sqsubseteq B_3$?



Statistics to control the logic?

- Use "**Google distance**" to decide which ones are reasonable to focus on
- Google distance
  - ≈ symmetric conditional probability of co-occurrence
  - ≈ estimate of semantic distance
  - ≈ estimate of "contribution" to $A \sqsubseteq B_1 \sqcap B_2 \sqcap B_3$



This work by Riste Gligorov

Statistics to define semantics?

- Many peers have many mappings
  on many terms
  to many other peers
- Mapping is good if results of
  "whispering game" are truthful
- Punish mappings that contribute to bad
  whispering results
- Network will converge
  to set of good mappings
  (or at least: consistent)

This work by
Karl Aberer

Statistics to define semantics?

- Meaning of terms =
  relations to other terms
- Determined by stochastic process
- Meaning ≈
  stable state of self-organising system
- **statistics = getting a system to a meaning-defining stable state**
- **logic = description of such a stable state**

- Note: meaning is still binary,
  classical truth-value
- Note: same system may have
  multiple stable states…

4 examples of “where does it break?”

- old assumptions that no longer hold,
- old approaches that no longer work

❶ Traditional complexity measures don’t work
- completeness, decidability, complexity

❷ Sometimes “hard in theory, easy in practice”
- Q/A over inconsistent ontologies is easy, but why?

❸ Meaning dependent on context
- meaning determined by background knowledge

❹ Logic versus statistics
- statistics in the logic
- statistics to control the logic
- statistics to determine semantics

Final comments

- These 4 “broken assumptions/old methods” were just examples. There are many more.
  (e.g. Hayes, Halpin on identity, equality and reference)

- Notice that they are interlinked, e.g
  ❷ hard theory/easy practice & ❶ complexity
  ❸ meaning in context & ❹ logic/statistics

- Working on these will not be SemWeb work per se, but
  - they will be inspired by SemWeb challenges
  - they will help the SemWeb effort (either V1 or V2)

Have fun with the puzzles!