From standard reasoning problems to non-standard reasoning problems and one step further

Uli Sattler
University of Manchester, University of Oslo
uli.sattler@manchester.ac.uk
Some Advertisement
Description Logics (DLs) have a long tradition in computer science and knowledge representation, being designed so that domain knowledge can be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.
Description Logics (DLs) have a long tradition in computer science and knowledge representation, being designed so that domain knowledge can be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.
Standard & Non-Standard Reasoning Problems
Standard Reasoning Problems

we all know them: given $C, D, O, T, A, \ldots$ decide/compute

- consistency/satisfiability
- subsumption
- classification
- query answering

- …all only involve entailment checks:

$$O \models \alpha$$

- …possibly many (classification!)
Non-Standard Reasoning Problems

we all know them: given $C, D, O, T, A, \ldots$

- $\text{msc}(a, O), \text{lcs}(C, D, O), \ldots$
- $\text{Justs}(\alpha, O), \text{PinPoint}(\alpha, O), \ldots$
- $\text{match}(C, P, O), \text{unify}(P_1, P_2, O), \ldots$
- $\text{x-mod}(\Sigma, O), \ldots$

- ...involve finding extreme $X$ such that ...
  - subset-minimal or
  - maximally/minimally strong
- ...possibly many such $X$s
Non-Standard Reasoning Problems

we all know them: given $C, D, \mathcal{O}, \mathcal{T}, \mathcal{A}, \ldots$
- $\text{msc}(a, \mathcal{O}), \text{lcs}(C, D, \mathcal{O}), \ldots$
- $\text{Justs}(\alpha, \mathcal{O}), \text{PinPoint} \ldots$
- $\text{match}(C, \mathcal{D}), \mathcal{O}, \ldots$
- $\text{x-mod}(\Sigma, \mathcal{O}), \ldots$
- ...involve finding extreme $X$ such that ...
  - subset-minimal or
  - maximally/minimally strong
- ...possibly many such $X$s

Are
- (conservative) rewritability
- (query) inseparability
also standard reasoning problems?
(Non-)Standard Reasoning: we know how to

understand problems:
• decidability & computational complexity
  – worst case
  – data
  – parametrised
  – …

understand solutions:
• soundness, completeness, termination
• relations between them
• complexity
  – see above
• practicability
  – worst case complexity ≠ best case complexity
  – amenable to optimisation
  – empirical evaluation
An interesting side note

from our empirical evaluation:
how many subsumption does classification involve?

<table>
<thead>
<tr>
<th>Ontology/Reasoner pairs</th>
<th>Fact</th>
<th>Hermit</th>
<th>JFact</th>
<th>Pellet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tests</td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

The number of tests for each Ontology/Reasoner pair is shown, with lines indicating the number of tests for different complexity classes: N*\log(N), N^2, and ST (Subsumption Test). The graphs show a comparison of the number of tests across different classes for each reasoner, illustrating how many subsumption tests are carried out in relation to a naive upper bound and an \(N\log(N)\) upper bound, ordered by \(N\) then number of names in \(O\) (y: log scale).

6.1.6 Discussion

A first answer to Question 1 is that (tableau) subsumption testing does not contribute at all to classification time for a substantial number of ontologies. We have established an empirical lower bound for BioPortal ontologies that do not involve subsumption testing at 46%. This lower bound is set because (1) there are most likely a number of ontologies that do not involve tests among the unsuccessfully classified ones and (2) only 33% of all ontology-reasoner pairs involved tests. The currently secured lower bound for ontologies actually requiring subsumption testing lies at 14% (i.e. the 47 out of 330 ontologies for which all four reasoners triggered a test). Note that, while this might seem like a very low number, these might be the 50 or so ontologies in the world that are hard and matter, and thus worth optimising for. As a side note, the low numbers of tests for HermiT and Pellet can perhaps be explained by their internal alternative deterministic engines (for example internal EL-reasoners), see Section 5.1.

It is quite interesting that only 10 out of those 146 ontologies that all reasoners processed caused at least one reasoner to fire a test—all of which are pure OWL 2 EL. Ontologies of the OWL 2 RL or OWL 2 QL family, or less expressive ontologies, did not cause any reasoner to actually fire a test. This suggests that for OWL 2 RL and OWL 2 QL ontologies at the very least, the application of modular techniques must be strictly motivated by a different argument than test avoidance or test easyfication. Another potentially interesting observation is that ontologies involving hard tests generally seem to contain rich role-level modelling, most prominently inverses and role hierarchies.

Subsumption test hardness rarely has a strong impact on classification performance. According to our threshold of “strong impact” at 40% of the overall classification time, FaCT++ encountered impactful ontologies 7.8% of the time, JFact 9.6% of the time, Pellet 4.2% of the time and HermiT only in 3 out of its 284 successes.

An interesting side note from our empirical evaluation: how many subsumption does classification involve?
Not always that straightforward

• Which problem/solution to consider when?
  • e.g., $\times$-mod$(\Sigma, \mathcal{O})$, ...
    • minimal/top/bottom/semantic/…
    • depends on size, signature, application, …
• but we know properties of/relations between solutions
  • smallest
  • self-contained
  • unique
  • depleting
  • …

• How to measure practicability?
  • benchmarks, ORE,..
Not always that straightforward

- Which problem/solution to consider when?
  - e.g., $x\text{-mod}(\Sigma, \mathcal{O}), \ldots$
    - minimal/top/bottom/semantic/…
    - depends on size, signature, application, …
  - but we know properties of/relations between solutions
    - smallest
    - self-contained
    - unique
    - depleting
    - …

- How to measure practicability?
  - benchmarks, ORE,..
Subjective Ontology-Based Problems
Subjective Ontology-Based Problems

are problems that are based on

- $C, D, O, T, A, \models, \ldots$ plus
- additional parameter(s)
Subjective Ontology-Based Problems

are problems that are based on

- \( C, D, \mathcal{O}, \mathcal{T}, A, \models, \ldots \) plus
- additional parameter(s)
- because objective solution is
  - not feasible/computable
  - or makes little sense
- e.g. in SROIQ: ComSubs(\( C, D, \{ C \sqsubseteq \forall R.(A \cap C), D \sqsubseteq \forall R.(A \cap D) \})
Subjective Ontology-Based Problems

are problems that are based on

• \( C, D, O, T, A, \models, \ldots \) plus
• additional parameter(s)
• because objective solution is
  • not feasible/computable
  • or makes little sense
• e.g. in SROIQ: ComSubs(\( C, D, \{ C \sqsubseteq \forall R.(A \cap C), \forall R.\forall R.\forall R.A, \ldots \})

\[= \{ \forall R.A, \forall R.\forall R.A, \forall R.\forall R.\forall R.A, \ldots \} \]
Subjective Ontology-Based Problems

are problems that are based on

- \( C, D, O, T, A, \models, \ldots \) plus
- additional parameter(s)

- because objective solution is
  - not feasible/computable
  - or makes little sense
- e.g. in \( SROIQ : \) ComSubs\( (C, D, \{ C \subseteq \forall R.(A \cap C), D \subseteq \forall R.(A \cap D)) \)  

or we want to capture quality criteria
- interestingness
- readability
- relevance \( \ldots \)
A subjective OB problem: Mining TBox Axioms from KBs or Finding Interesting Correlations
Mining TBox axioms from KBs

- learn (implicit) correlations in our data
- get interesting insights into domain

Do not confuse with
(exact) learning of TBoxes (via probing queries)
Mining TBox axioms from KBs

- Correlations in KB = classical machine learning
  - automatic generation of knowledge from data
    - taking **background knowledge** in KB into account
    - unbiased: let the data speak!
    - unsupervised (no positive/negative examples)
    - **Semantic Data Mining**
Mining TBox axioms from KBs

- Correlations in KB = classical machine learning
  - automatic generation of knowledge from data
    - taking background knowledge in KB into account
    - unbiased: let the data speak!
    - unsupervised (no positive/negative examples)
    - *Semantic Data Mining*
Mining TBox axioms from KBs

- Which kind of hypotheses to capture correlations in KB?
  1. expressive: GCIs, role inclusions
  2. readable
  3. logically sound
  4. statistically sound
2. Readable Hypotheses

• A hypothesis is
  – a small set of short axioms
    • fewer than $n_{\text{max}}$ axioms
    • with concepts shorter than $\ell_{\text{max}}$
  – in a suitable DL: $\text{ALCHI} \ldots \text{SROIQ}$
  – free of redundancy
    • no superfluous parts
    ✓ preferred laconic justifications
3. Logically Sound Hypotheses

• A hypothesis $H$ should be

✓ informative: $\forall \alpha \in H : \mathcal{O} \not\models \alpha$
  ✓ we want to mine new axioms

✓ consistent: $\mathcal{O} \cup H \not\models \top \sqsubseteq \bot$

✓ non-redundant among all hypotheses:
  • there is no $H', H \in \mathcal{H} : H \neq H'$ and $H' \equiv H$
3. Logically Sound Hypotheses

- A hypothesis $H$ should be
  - ✓ informative: $\forall \alpha \in H : \emptyset \not\models \alpha$
    - ✓ we want to mine new axioms
  - ✓ consistent: $\emptyset \cup H \not\models T \subseteq \bot$
  - ✓ non-redundant among all hypotheses:
    - there is no $H', H \in \mathcal{H} : H \neq H'$ and $H' \equiv H$

- Different hypotheses can be compared wrt. their
  - ✓ logical strength:
    - ? maximally strong?
      - no: overfitting!
    - ? minimally strong?
      - no: under-fitting
  - ✓ reconciliatory power
    - • brings together terms so far only loosely related
4. Statistically Sound Hypotheses

- we need to assess *data support* of hypothesis
- introduce metrics that capture *quality* of an axiom
  - learn from association rule mining (ARM):
    - count individuals that *support* a GCI
    - count instances, neg instances, non-instances
      - using standard DL semantics, OWA, TBox, entailments,….
    - no ‘artificial closure’
  - make sure you treat a GCI as an *axiom* and not as a *rule*
    - contrapositive!
  - coverage, support, …., lift
4. Statistically Sound Hypotheses

Some useful notation:

- \( \text{Inst}(C, \mathcal{O}) := \{ a \mid \mathcal{O} \models C(a) \} \)
- \( \text{UnKn}(C, \mathcal{O}) := \text{Inst}(\top, \mathcal{O}) \setminus (\text{Inst}(C, \mathcal{O}) \cup \text{Inst}(\neg C, \mathcal{O})) \)
- relativized: \( P(C, \mathcal{O}) := \# \text{Inst}(C, \mathcal{O}) / \# \text{Inst}(\top, \mathcal{O}) \)
- projection tables:

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind2</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>Ind3</td>
<td>?</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind4</td>
<td>?</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Axioms

some axiom measures easily adapted from ARM:
for a GCI $C \subseteq D$ define its metrics as follows:

<table>
<thead>
<tr>
<th>Coverage</th>
<th>basic</th>
<th>relativized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$# \text{Inst}(C, O)$</td>
<td>$P(C, O)$</td>
</tr>
<tr>
<td>Support</td>
<td>$# \text{Inst}(C \cap D, O)$</td>
<td>$P(C \cap D, O)$</td>
</tr>
<tr>
<td>Contradiction</td>
<td>$# \text{Inst}(C \cap \neg D, O)$</td>
<td>$P(C \cap \neg D, O)$</td>
</tr>
<tr>
<td>Assumption</td>
<td>$# \text{Inst}(C, O) \cap \text{UnKn}(D, O)$</td>
<td>...</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, O)/P(C, O)$</td>
<td></td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, O)/P(C, O)P(D, O)$</td>
<td></td>
</tr>
</tbody>
</table>

where $P(X, O) = \# \text{Ind}(X, O)/\# \text{Ind}(\top, O)$
4. Statistically Sound Hypotheses: Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>relativized</th>
<th>$A \sqsubseteq B$</th>
<th>$B \sqsubseteq C_1$</th>
<th>$B \sqsubseteq C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>$P(C, O)$</td>
<td>200/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Support</td>
<td>$P(C \cap D, O)$</td>
<td>180/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Assumption</td>
<td>...</td>
<td>20/400</td>
<td>0</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, O)/P(C, O)$</td>
<td>180/200</td>
<td>180/180</td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, O)/P(C, O)P(D, O)$</td>
<td>400/200</td>
<td>400/200</td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>relativized</th>
<th>$A \subseteq B$</th>
<th>$B \subseteq C_1$</th>
<th>$B \subseteq C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>$P(C, O)$</td>
<td>200/400</td>
<td>180/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Support</td>
<td>$P(C \cap D, O)$</td>
<td>180/400</td>
<td>180/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Assumption</td>
<td>...</td>
<td>20/400</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, O)/P(C, O)$</td>
<td>180/200</td>
<td>180/180</td>
<td>180/180</td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, O)/P(C, O)P(D, O)$</td>
<td>400/200</td>
<td>400/200</td>
<td>400/180</td>
</tr>
</tbody>
</table>
### 4. Statistically Sound Hypotheses: Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>relativized</th>
<th>(A \subseteq B)</th>
<th>(B \subseteq C1)</th>
<th>(B \subseteq C2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>(P(C, O))</td>
<td>200/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Support</td>
<td>(P(C \cap D, O))</td>
<td>180/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Assumption</td>
<td>...</td>
<td>20/400</td>
<td>0</td>
</tr>
<tr>
<td>Confidence</td>
<td>(P(C \cap D, O)/P(C, O))</td>
<td>180/200</td>
<td>180/180</td>
</tr>
<tr>
<td>Lift</td>
<td>(P(C \cap D, O)/P(C, O)P(D, O))</td>
<td>400/200</td>
<td>400/200</td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
</table>
| Ind1 | X | X | X | X | ...
| ... | ... | ... | ... | ... | ...
| Ind180 | X | X | X | X | ...
| Ind181 | X | ? | X | ? | ...
| ... | ... | ... | ... | ... | ...
| Ind200 | X | ? | X | ? | ...
| Ind201 | ? | ? | ? | ? | ...
| ... | ... | ... | ... | ... | ...
| Ind400 | ? | ? | ? | ? | ...

<table>
<thead>
<tr>
<th>relativized</th>
<th>( A \subseteq B )</th>
<th>( B \subseteq C_1 )</th>
<th>( B \subseteq C_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>( P(C, O) )</td>
<td>200/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Support</td>
<td>( P(C \cap D, O) )</td>
<td>180/400</td>
<td>180/400</td>
</tr>
<tr>
<td>Assumption</td>
<td>...</td>
<td>20/400</td>
<td>0</td>
</tr>
<tr>
<td>Confidence</td>
<td>( P(C \cap D, O)/P(C, O) )</td>
<td>180/200</td>
<td>180/180</td>
</tr>
<tr>
<td>Lift</td>
<td>( P(C \cap D, O)/P(C, O)P(D, O) )</td>
<td>400/200</td>
<td>400/200</td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Example

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>relativized</th>
<th>$A \sqsubseteq B$</th>
<th>$B \sqsubseteq C1$</th>
<th>$B \sqsubseteq C2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>$P(C, \mathcal{O})$</td>
<td>0.5</td>
<td>0.45</td>
</tr>
<tr>
<td>Support</td>
<td>$P(C \cap D, \mathcal{O})$</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Assumption</td>
<td>$P(C \cap D, \mathcal{O})/P(C, \mathcal{O})$</td>
<td>0.05</td>
<td>0</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$</td>
<td>0.45</td>
<td>1</td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!

- so: turn each GCI $X \subseteq Y$ into equivalent $X \cup \neg Y \subseteq Y \cup \neg X$

read $C$ below as ‘the resulting LHS’…
read $D$ below as ‘the resulting RHS’…

<table>
<thead>
<tr>
<th>Coverage</th>
<th>main</th>
<th>relativized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Inst$(C, O)$</td>
<td>$P(C, O)$</td>
</tr>
<tr>
<td>Support</td>
<td># Inst$(C \cap D, O)$</td>
<td>$P(C \cap D, O)$</td>
</tr>
<tr>
<td>Contradiction</td>
<td># Inst$(C \cap \neg D, O)$</td>
<td>$P(C \cap \neg D, O)$</td>
</tr>
<tr>
<td>Assumption</td>
<td># Inst$(C, O) \cap \text{UnKn}(D, O)$</td>
<td>...</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, O)/P(C, O)$</td>
<td>...</td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, O)/P(C, O)P(D, O)$</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!
- so: turn each GCI $X \subseteq Y$ into equivalent $\neg Y \subseteq \neg X$
  
<table>
<thead>
<tr>
<th>Coverage</th>
<th># Inst($C, O$)</th>
<th>$P(C, O)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td># Inst($C \cap D, O$)</td>
<td>$P(C \cap D, O)$</td>
</tr>
<tr>
<td>Contradiction</td>
<td># Inst($C \cap \neg D, O$)</td>
<td>$P(C \cap \neg D, O)$</td>
</tr>
<tr>
<td>Assumption</td>
<td># Inst($C, O \cap \text{UnKn}(D, O)$)</td>
<td>...</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, O) / P(C, O)$</td>
<td></td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, O) / P(C, O) P(D, O)$</td>
<td></td>
</tr>
</tbody>
</table>

Axiom measures are semantically faithful, i.e., $\text{Ass}(A \subseteq B, O) = \text{Ass}(\neg B \subseteq \neg A, O)$
4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!
- so: turn each GCI $X \subseteq Y$ into
  read $\bar{C}$ below

Axiom measures are not semantically faithful, e.g.,

$$\text{Support}(A \subseteq B, \mathcal{O}) \neq \text{Support}(\top \subseteq \neg A \sqcup B, \mathcal{O})$$

<table>
<thead>
<tr>
<th>Coverage</th>
<th>$#\text{Inst}(C \cap D, \mathcal{O})$</th>
<th>$P(C, \mathcal{O})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support</td>
<td>$#\text{Inst}(C \cap D, \mathcal{O})$</td>
<td>$P(C \cap D, \mathcal{O})$</td>
</tr>
<tr>
<td>Contradiction</td>
<td>$#\text{Inst}(C \cap \neg D, \mathcal{O})$</td>
<td>$P(C \cap \neg D, \mathcal{O})$</td>
</tr>
<tr>
<td>Assumption</td>
<td>$#\text{Inst}(C, \mathcal{O}) \cap \text{UnKn}(D, \mathcal{O})$</td>
<td>...</td>
</tr>
<tr>
<td>Confidence</td>
<td>$P(C \cap D, \mathcal{O}) / P(C, \mathcal{O})$</td>
<td></td>
</tr>
<tr>
<td>Lift</td>
<td>$P(C \cap D, \mathcal{O}) / P(C, \mathcal{O}) P(D, \mathcal{O})$</td>
<td></td>
</tr>
</tbody>
</table>

Goal: mine small sets of (short) axioms

- more readable
  - close to what people write
- synergy between axioms should lead to better quality
- how to measure their qualities?
Goal: learn small sets of (short) axioms

• more readable
  - close to what people write
• synergy between axioms should lead to better quality
• how to measure their qualities?
  • …easy:
    1. rewrite set into single axiom as usual
    2. measure resulting axiom


\[ H_1 = \{ A \sqsubseteq B, B \sqsubseteq C_1 \} \]
\[ \equiv \{ \top \sqsubseteq (\neg A \cup B) \cap (\neg B \cup C_1) \} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Coverage</th>
<th>Support</th>
<th>Assumption</th>
<th>Confidence</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.45</td>
<td>0.05</td>
<td>0.45</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ H1 = \{ A \subseteq B, B \subseteq C1 \} \]
\[ \equiv \{ \top \subseteq (\neg A \cup B) \cap (\neg B \cup C1) \} \]

<table>
<thead>
<tr>
<th></th>
<th>A \subseteq B</th>
<th>B \subseteq C1</th>
<th>B \subseteq C2</th>
<th>H1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>0.5</td>
<td>0.45</td>
<td>0.45</td>
<td>1</td>
<td>always!</td>
</tr>
<tr>
<td>Support</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>min</td>
</tr>
<tr>
<td>Assumption</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.55</td>
<td>?</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td>0.45</td>
<td>support!</td>
</tr>
<tr>
<td>Lift</td>
<td>2</td>
<td>2</td>
<td>2.22</td>
<td>1</td>
<td>always!</td>
</tr>
</tbody>
</table>

Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- synergy between axioms should lead to better quality
- how to measure their qualities?
  - sum/average quality of their axioms!

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( A \subseteq B )</th>
<th>( B \subseteq C_1 )</th>
<th>( H_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>0.5</td>
<td>0.45</td>
<td>0.475?</td>
</tr>
<tr>
<td>Support</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Assumption</td>
<td>0.05</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.45</td>
<td>1</td>
<td>?</td>
</tr>
<tr>
<td>Lift</td>
<td>2</td>
<td>2</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

<table>
<thead>
<tr>
<th></th>
<th>A \subseteq B</th>
<th>B \subseteq C_1</th>
<th>H1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>0.475?</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td></td>
</tr>
</tbody>
</table>

| Coverage | 0.5 | 0.45 |
| Support  | 0.45 | 0.45 |
| Assumption | 0.05 | 0.05 |
| Confidence | 0.45 | 1.0  |
| Lift     | 2    | 2    |

\[ H1 = \{ A \subseteq B, B \subseteq C1\} \]

\[ H2 = \{ A \subseteq B, B \subseteq C2\} \]

<table>
<thead>
<tr>
<th></th>
<th>A ⊆ B</th>
<th>B ⊆ C1</th>
<th>B ⊆ C2</th>
<th>H1</th>
<th>H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>0.5</td>
<td>0.45</td>
<td>0.45</td>
<td>0.475?</td>
<td>0.475?</td>
</tr>
<tr>
<td>Support</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>Assumption</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Lift</td>
<td>2</td>
<td>2</td>
<td>2.22</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ H_1 = \{ A \sqsubseteq B, B \sqsubseteq C_1 \} \]
\[ H_2 = \{ A \sqsubseteq B, B \sqsubseteq C_2 \} \]

<table>
<thead>
<tr>
<th></th>
<th>A ⊆ B</th>
<th>B ⊆ C1</th>
<th>B ⊆ C2</th>
<th>H1</th>
<th>H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>...</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
</tbody>
</table>

| Coverage | 0.5 | 0.45 | 0.45 | 0.475? | 0.475? |
| Support  | 0.45| 0.45 | 0.45 | 0.45   | 0.45   |
| Assumption | 0.05 | 0 | 0 | 0.05 | 0.05 |
| Confidence| 0.45| 1 | 1 | ? | ? |
| Lift     | 2   | 2 | 2.22 | ? | ? |

Goal: learn small sets of (short) axioms

• more readable
  - close to what people write
• synergy between axioms should lead to better quality
• how to measure their qualities?
  • observe that a good hypothesis
    • allows us to shrink our ABox since it
    • captures recurring patterns
  • (minimum description length induction)
Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?
  - observe that a good hypothesis
    - allows us to *shrink* our ABox since it
    - captures *recurring patterns*
  - use this shrinkage factor to measure a hypothesis’
    - fitness - support by data
    - braveness - number of assumptions
Capturing shrinkage...for fitness

• Fix a finite set of
  – concepts $\mathbb{C}$, closed under negation
  – roles $\mathbb{R}$
Capturing shrinkage...for fitness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a projection:
  $$\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \{ C(a) \mid \mathcal{O} \models C(a) \land C \in \mathbb{C} \} \cup \{ R(a, b) \mid \mathcal{O} \models C(a) \land R \in \mathbb{R} \}$$
Capturing shrinkage…for fitness

- Fix a finite set of
  - concepts $C$, closed under negation
  - roles $R$

- Define a projection:
  $$\pi(O, C, R) = \{ C(a) \mid O \models C(a) \land C \in C \} \cup \{ R(a, b) \mid O \models C(a) \land R \in R \}$$

- For an ABox, define its description length:
  $$d\text{Len}(A, O) = \min\{ \ell(A') \mid A' \cup O \equiv A \cup O \}$$
Capturing shrinkage...for fitness

- Fix a finite set of
  - concepts \( \mathbb{C} \), closed under negation
  - roles \( \mathbb{R} \)
- Define a projection:
  \[
  \pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \{ C(a) \mid \mathcal{O} \models C(a) \land C \in \mathbb{C} \} \cup \{ R(a, b) \mid \mathcal{O} \models C(a) \land R \in \mathbb{R} \}
  \]
- For an ABox, define its description length:
  \[
  d\text{Len}(\mathcal{A}, \mathcal{O}) = \min\{ \ell(\mathcal{A}') \mid \mathcal{A}' \cup \mathcal{O} \equiv \mathcal{A} \cup \mathcal{O} \}
  \]
- Define the fitness of a hypothesis \( H \):
  \[
  \text{fitn}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) = d\text{Len}(\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}), \mathcal{T}) - d\text{Len}(\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}), \mathcal{T} \cup H)
  \]
Capturing shrinkage…for braveness

• Fix a finite set of
  – concepts \( C \), closed under negation
  – roles \( R \)

• Define a projection:
  \[
  \pi(\mathcal{O}, C, R) = \{ C(a) \mid \mathcal{O} \models C(a) \land C \in C \} \cup \{ R(a, b) \mid \mathcal{O} \models C(a) \land R \in R \}
  \]
Capturing shrinkage...for braveness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a projection:
  \[
  \pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \{ C(a) \mid \mathcal{O} \models C(a) \land C \in \mathbb{C} \} \cup \{ R(a, b) \mid \mathcal{O} \models C(a) \land R \in \mathbb{R} \}
  \]
- Define a hypothesis’ assumptions:
  \[
  \text{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = \pi(\mathcal{O} \cup H, \mathbb{C}, \mathbb{R}) \setminus \pi(\mathcal{O}, \mathbb{C}, \mathbb{R})
  \]
Capturing shrinkage...for braveness

• Fix a finite set of
  – concepts \( C \), closed under negation
  – roles \( R \)

• Define a projection:
  \[
  \pi(O, C, R) = \{ C(a) \mid O \models C(a) \land C \in C \} \cup \\
  \{ R(a, b) \mid O \models C(a) \land R \in R \}
  \]

• Define a hypothesis’ assumptions:
  \[
  \text{Ass}(O, H, C, R) = \pi(O \cup H, C, R) \setminus \pi(O, C, R)
  \]

• Define the braveness of a hypothesis \( H \):
  \[
  \text{brave}(H, O, C, R) = \text{dLen}(\text{Ass}(O, H, C, R), O)
  \]
Capturing shrinkage...for braveness

- Fix a finite set of
  - concepts \( C \), closed under negation
  - roles \( R \)
- Define a projection:
  \[ \pi(O, C, R) \]

Axiom set measures are semantically faithful, i.e.,

\[
H \equiv H' \Rightarrow \quad \text{fitn}(H, O, C, R) = \text{fitn}(H', O, C, R)
\]

\[
\text{brave}(H, O, C, R) = \text{brave}(H', O, C, R)
\]

\[
(O \cup H, C, R) \setminus \pi(O, C, R)
\]

- Define the braveness of a hypothesis \( H \):
  \[
  \text{brave}(H, O, C, R) = \text{dLen}(\text{Ass}(O, H, C, R), O)
  \]

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\text{fitn}(H_1, A, \ldots) = d\text{Len}(\pi(A, \ldots), \emptyset) - d\text{Len}(\pi(A, \ldots), H_1) = 760 - 380 = 380
\]

\[
\text{brave}(H_1, A, \ldots) = d\text{Len}(\text{Ass}(A, H_1, \ldots), A) = 20
\]

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Ind1} & X & X & X & X \\
\hline
\text{Ind180} & X & X & X & X \\
\hline
\text{Ind181} & X & ? & X & ? \\
\hline
\text{Ind200} & X & ? & X & ? \\
\hline
\text{Ind201} & ? & ? & ? & ? \\
\hline
\text{Ind400} & ? & ? & ? & ? \\
\hline
\end{array}
\]

\[
\text{fitn}(H_1, A, \ldots) = d\text{Len}(\pi(A, \ldots), \emptyset) - d\text{Len}(\pi(A, \ldots), H_1) = 760 - 380 = 380
\]

\[
\text{brave}(H_1, A, \ldots) = d\text{Len}(\text{Ass}(A, H_1, \ldots), A) = 20
\]

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

\[ H_2 = \{ A \subseteq B, B \subseteq C_2 \} \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\text{fitn}(H_1, A, \ldots) = \]
\[
dLen(\pi(A, \ldots), \emptyset) - dLen(\pi(A, \ldots), H_1) = 760 - 380 = 380
\]

\[
\text{fitn}(H_2, A, \ldots) = \]
\[
dLen(\pi(A, \ldots), \emptyset) - dLen(\pi(A, \ldots), H_2) = 760 - 400 = 360
\]

\[
\text{brave}(H_1, A, \ldots) = dLen(\text{Ass}(A, H_1, \ldots), A) = 20
\]

\[
\text{brave}(H_2, A, \ldots) = dLen(\text{Ass}(A, H_2, \ldots), A) = 40
\]

\[ H_1 = \{ A \subseteq B, B \subseteq C_1 \} \]

\[ H_2 = \{ A \subseteq B, B \subseteq C_2 \} \]

\[ H_1 \gg H_2 \]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ind181</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind200</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>?</td>
</tr>
<tr>
<td>Ind201</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Ind400</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

\[
\text{fitn}(H_1, A, \ldots) = \\
d\text{Len}(\pi(A, \ldots), \emptyset) - d\text{Len}(\pi(A, \ldots), H_1) = 760 - 380 = 380
\]

\[
\text{fitn}(H_2, A, \ldots) = \\
d\text{Len}(\pi(A, \ldots), \emptyset) - d\text{Len}(\pi(A, \ldots), H_2) = 760 - 400 = 360
\]

\[
\text{brave}(H_1, A, \ldots) = d\text{Len}(\text{Ass}(A, H_1, \ldots), A) = 20
\]

\[
\text{brave}(H_2, A, \ldots) = d\text{Len}(\text{Ass}(A, H_2, \ldots), A) = 40
\]

Example: empty TBox, ABox $\mathcal{A}$

$$\text{fitn}(\{ X \sqsubseteq \forall R.A \}, \mathcal{A}, \ldots) = d\text{Len}(\pi(\mathcal{A}, \ldots), \emptyset) - d\text{Len}(\pi(\mathcal{A}, \ldots), \{ X \sqsubseteq \forall R.A \}) = 12 - 9 = 3$$

$$\text{brave}(\{ X \sqsubseteq \forall R.A \}, \mathcal{A}, \ldots) = d\text{Len}(\text{Ass}(\mathcal{A}, \{ X \sqsubseteq \forall R.A \}, \ldots), \mathcal{A}) = 1$$

Example: empty TBox, ABox $\mathcal{A}$

\[
\text{fitn}\left(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots\right) = \text{dLen}\left(\pi(\mathcal{A}, \ldots), \emptyset\right) - \\
\text{dLen}\left(\pi(\mathcal{A}, \ldots), \{X \sqsubseteq \forall R.A\}\right) \\
= 12 - 9 \\
= 3
\]

\[
\text{brave}\left(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots\right) = \text{dLen}\left(\text{Ass}(\mathcal{A}, \{X \sqsubseteq \forall R.A\}, \ldots), \mathcal{A}\right) \\
= 1
\]

Example: empty TBox, ABox $\mathcal{A}$

$$\text{fitn}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) = d\text{Len}(\pi(\mathcal{A}, \ldots), \emptyset) - d\text{Len}(\pi(\mathcal{A}, \ldots), \{X \sqsubseteq \forall R.A\})$$
$$= 12 - 9$$
$$= 3$$

$$\text{brave}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) = d\text{Len}(\text{Ass}(\mathcal{A}, \{X \sqsubseteq \forall R.A\}, \ldots), \mathcal{A})$$
$$= 1$$
phew...
Remember:

we wanted to mine axioms!
So, what have we got?

- (Sets of) axioms as Hypotheses
- Loads of measures to capture
  1. axiom hypothesis’ coverage, support, assumption, lift, …
  2. set of axioms hypothesis fitness, braveness
    - with a focus of a concept/role spaces $C, R$

- Can we compute these measures?
  - easy for (1), tricky for (2): 

So, what have we got?
So, what have we got?

- (Sets of) axioms as Hypotheses
- Loads of measures to capture
  1. axiom hypothesis’ coverage, support, assumption, lift, …
  2. set of axioms hypothesis fitness, braveness
    - with a focus of a concept/role spaces $\mathbb{C}, \mathbb{R}$
- What are their properties?
  - semantically faithful:
    \[
    \mathcal{O} \models H \Rightarrow \text{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = 0
    \]
    \[
    H \equiv H' \Rightarrow \text{fitn}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = \text{fitn}(\mathcal{O}, H', \mathbb{C}, \mathbb{R})
    \]
    ...
- Can we compute these measure?
  - easy for (1), tricky for (2): $d\text{Len}(A, \mathcal{O}) = \min\{\ell(A') \mid A' \cup \mathcal{O} \equiv A \cup \mathcal{O}\}$
So, what have we got? (2)

• If we can compute measure, how feasible is this?
• If “feasible”,
  – do these measures correlate?
  – how independent are they?
• For which DLs & inputs can we create & evaluate hypotheses?
• Which measures indicate interesting hypothesis?
• What is the shape for interesting hypothesis?
  – are longer/bigger hypotheses better?
• What do we do with them?
  – how do we guide users through these?
Slava implements: DL Miner

4.2. DESIGNING DL-MINER

Ontology Cleaner

Hypothesis Constructor

Hypothesis Evaluator

Hypothesis Sorter

$\mathcal{L}, \Sigma$ parameters

$Q$ parameters

$\text{axiom}(s) m_1,m_2,m_3,…$

$rf(H)$

$\mathcal{H}$

$\text{domain expert and ontology engineers are supposed to navigate through the hypotheses using the quality and ranking functions. Thus, all hypotheses can be methodically examined. Clearly, it is possible to select only best hypotheses if necessary. As the reader will find in the following, hypotheses of DL-Miner can, in fact, be used for various purposes and in different scenarios.}$

In the following, we clarify the parameters and unfold the functionality of each block. Hypothesis Evaluator is covered in Chapter 5, where we define quality measures that can be used in $Q$, and Chapter 6, where we develop techniques to compute those measures. Hypothesis Constructor is explained in Chapter 7 where we show how to construct suitable concepts $C$ (roles $R$) given an language bias $L$ and generate hypotheses $H$ from $C(R)$. Ontology Cleaner and Hypothesis Sorter are both covered in Chapter 8 where we also integrate all techniques in DL-Miner. Finally, we empirically evaluate DL-Miner in Chapter 9.
Slava implements: DL Miner

Figure 4.1: Architecture of DL-Miner

- Hypothesis Sorter, given the equality function $qf(\cdot)$, orders hypotheses $H$ according to the binary relation $O$. The result is the ranking function $rf(H)$ that returns the quality rank of a hypothesis $H$. The output of DL-Miner is a set $H$ of hypotheses, quality function $qf(\cdot)$, and ranking function $rf(\cdot)$. Domain experts and ontology engineers are supposed to navigate through the hypotheses using the quality and ranking functions. Thus, all hypotheses can be methodically examined. Clearly, it is possible to select only the best hypotheses if necessary. As the reader will find in the following, hypotheses of DL-Miner can, in fact, be used for various purposes and in different scenarios.

In the following, we clarify the parameters and unfold the functionality of each block. Hypothesis Evaluator is covered in Chapter 5, where we define quality measures that can be used in $Q$, and Chapter 6, where we develop techniques to compute those measures. Hypothesis Constructor is explained in Chapter 7 where we show how to construct suitable concepts $C(R)$ given an language bias $L$ and generate hypotheses $H$ from $C(R)$. Ontology Cleaner and Hypothesis Sorter are both covered in Chapter 8 where we also integrate all techniques in DL-Miner. Finally, we empirically evaluate DL-Miner in Chapter 9.

3 The symbol "\cdot" stands for the argument so that the function is recognized irrelevant.

4 When $O$ and $Q$ are clear from the context, we denote the binary relation $O, Q$ by $\cdot$. 
DL Miner: Hypothesis Constructor

Easy:

- construct all concepts $C1, C2, \ldots$
  - finitely many thanks to language bias $\mathcal{L}$
- check for each $C_i \sqsubseteq C_j$ whether it’s logically ok:
  - $\mathcal{O} \cup \{C_i \sqsubseteq C_j\} \not\models T \sqsubseteq \bot$
  - $\mathcal{O} \not\models C_i \sqsubseteq C_j$

if yes, add it to $\mathcal{H}$

- remove redundant hypotheses from $\mathcal{H}$
Easy:

- construct all concepts $C_1, C_2, \ldots$
  - finitely many thanks to language bias
- check for each
  - if yes, add it to
- remove redundant hypotheses from $H$

DL Miner: Hypothesis Constructor

**Bonkers!**

Even for $\mathcal{EL}$,

- 100 concept/role names
- 4 max length of concepts $C_i$
- $\sim 100,000,000$ concepts $C_i$
- $\sim 100,000,000^2$ GCIs to test
DL Miner: Hypothesis Constructor

Easy:

- construct all concepts $C_1, C_2, \ldots$
  - finitely many thanks to language bias
- check for each $C_i$ whether it's logically ok:
  - $O \cup \exists \neq \emptyset$
  - $O \neq \emptyset$
  - if yes, add it to $H$
- remove redundant hypotheses from $H$

Bonkers!

Even for $\mathcal{EL}$,

- $n$ concept/role names
- $k$ max length of concepts $C_i$
- $n^k$ concepts $C_i$
- $n^{2k}$ GCIs to test
DL Miner: Hypothesis Constructor

Use a refinement operator to build $C_i$ informed by ABox

- used in concept learning, conceptual blending

- Given a logic $\mathcal{L}$, define a refinement operator as
  - a function $\rho : \text{Conc}(\mathcal{L}) \mapsto \mathcal{P}(\text{Conc}(\mathcal{L}))$ such that, for each $C \in \mathcal{L}, C' \in \rho(C) : C' \subseteq C$

- A refinement operator is
  - proper if, for all $C \in \mathcal{L}, C' \in \rho(C) : C' \not= C$
  - complete if, for all $C, C' \in \mathcal{L} : C' \subseteq C$
    then there is some $n, C'' \equiv C$
    with $C' \in \rho^n(C'')$

- suitable if, for all $C \in \mathcal{L}$ there is $n, C' \in \rho^n(\top) : C' \equiv C$ and
  $\ell(C') \leq \ell(C)$
Use a refinement operator to build $Ci$ informed by $ABox$ – used in concept learning, conceptual blending

• Given a logic $\mathcal{L}$, define a refinement operator as:
  – a function $\rho : \text{Conc}(\mathcal{L}) \rightarrow 2^{\mathcal{D}(\mathcal{L})}$
    for each $C \in \text{Conc}(\mathcal{L})$
• A refinement operator is:
  – proper if, for all $C \in \text{Conc}(\mathcal{L})$:
    $C \nsubseteq C$
  – complete if, for all $C \in \text{Conc}(\mathcal{L})$:
    $C \nsubseteq C$
  – suitable if, for all $C \in \mathcal{L}$ there is $n, C' \in \rho^n(\top) : C' \equiv C$ and $\ell(C') \leq \ell(C)$

Great: there are known refinement operators (proper, complete, suitable, …) for $\text{ALC}$ [LehmHitzler2010] for $ALC$ [LehmHitzler2010]
DL Miner: Concept Constructor

Algorithm 8 DL-Apriori \((\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{\text{max}}, p_{\text{min}})\)

1: inputs
2: \(\mathcal{O} := \mathcal{T} \cup \mathcal{A}\): an ontology
3: \(\Sigma\): a finite set of terms such that \(\mathcal{T} \in \Sigma\)
4: \(\mathcal{DL}\): a DL for concepts
5: \(\ell_{\text{max}}\): a maximal length of a concept such that \(1 \leq \ell_{\text{max}} < \infty\)
6: \(p_{\text{min}}\): a minimal concept support such that \(0 < p_{\text{min}} \leq |\text{in}(\mathcal{O})|\)
7: outputs
8: \(\mathcal{C}\): the set of suitable concepts
9: do
10: \(\mathcal{C} \leftarrow \emptyset\) % initialise the final set of suitable concepts
11: \(\mathcal{D} \leftarrow \{\mathcal{T}\}\) % initialise the set of concepts yet to be specialised
12: \(\rho \leftarrow \text{getOperator}(\mathcal{DL})\) % initialise a suitable operator \(\rho\) for \(\mathcal{DL}\)
13: while \(\mathcal{D} \neq \emptyset\) do
14: \(C \leftarrow \text{pick}(\mathcal{D})\) % pick a concept \(C\) to be specialised
15: \(\mathcal{D} \leftarrow \mathcal{D}\backslash\{C\}\) % remove \(C\) from the concepts to be specialised
16: \(\mathcal{C} \leftarrow \mathcal{C}\cup\{C\}\) % add \(C\) to the final set
17: \(\rho_C \leftarrow \text{specialise}(C, \rho, \Sigma, \ell_{\text{max}})\) % specialise \(C\) using \(\rho\)
18: \(\mathcal{D}_C \leftarrow \{D \in \text{urc}(\rho_C) \mid \exists D' \in \mathcal{C} \cup \mathcal{D} : D' \equiv D\}\) % discard variations
19: \(\mathcal{D} \leftarrow \mathcal{D} \cup \{D \in \mathcal{D}_C \mid p(D, \mathcal{O}) \geq p_{\text{min}}\}\) % add suitable specialisations
20: end while
21: return \(\mathcal{C}\)
DL Miner: Concept Constructor

Algorithm 8 DL-Apriori ($\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{\text{max}}, p_{\text{min}}$)

1: inputs
2: $\mathcal{O} := \mathcal{T} \cup \mathcal{A}$: an ontology
3: $\Sigma$: a finite set of terms such that $\top \in \Sigma$
4: $\mathcal{DL}$: a DL for concepts
5: $\ell_{\text{max}}$: a maximal length of a concept such that $1 \leq \ell_{\text{max}} < \infty$
6: $p_{\text{min}}$: a minimal concept support such that $0 < p_{\text{min}} \leq |\text{in}(\mathcal{O})|$
7: outputs
8: $\mathcal{C}$: the set of suitable concepts
9: do
10: $\mathcal{C} \leftarrow \emptyset$ % initialise the final set of suitable concepts
11: $\mathcal{D} \leftarrow \{\top\}$ % initialise the set of concepts yet to be specialised
12: $\rho \leftarrow \text{getOperator}(\mathcal{DL})$ % initialise a suitable operator $\rho$ for $\mathcal{DL}$
13: while $\mathcal{D} \neq \emptyset$ do
14: $C \leftarrow \text{pick}(\mathcal{D})$ % pick a concept $C$ to be specialised
15: $\mathcal{D} \leftarrow \mathcal{D}\setminus\{C\}$ % remove $C$ from the concepts to be specialised
16: $\mathcal{C} \leftarrow \mathcal{C} \cup \{C\}$ % add $C$ to the final set
17: $\rho_C \leftarrow \text{specialise}(C, \rho, \Sigma, \ell_{\text{max}})$ % specialise $C$ using $\rho$
18: $\mathcal{D}_C \leftarrow \{D \in \text{urc}(\rho_C) \mid \exists D' \in \mathcal{C} \cup \mathcal{D} : D' \equiv D\}$ % discard variations
19: $\mathcal{D} \leftarrow \mathcal{D} \cup \{D \in \mathcal{D}_C \mid p(D, \mathcal{O}) \geq p_{\text{min}}\}$ % add suitable specialisations
20: end while
21: return $\mathcal{C}$

specialise concepts only if they have $\geq \ell_{\text{max}}$ instances!
**Algorithm 8 DL-APRIORI** \((\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{max}, p_{min})\)

1: inputs
2: \(\mathcal{O} := \mathcal{T} \cup \mathcal{A}\): an ontology
3: \(\Sigma\): a finite set of terms such that \(\mathcal{T} \in \Sigma\)
4: \(\mathcal{DL}\): a DL for concepts
5: \(\ell_{max}\): a maximal length of a concept such that \(1 \leq \ell_{max} < \infty\)
6: \(p_{min}\): a minimal concept support such that \(0 < p_{min} \leq |\text{in} (\mathcal{O})|\)
7: outputs
8: \(\mathcal{C}\): the set of suitable concepts
9: do
10: \(\mathcal{C} \leftarrow \emptyset\) % initialise the final set of suitable concepts
11: \(\mathcal{D} \leftarrow \{\top\}\) % initialise the set of concepts yet to be specialised
12: \(\rho \leftarrow \text{getOperator}(\mathcal{DL})\) % initialise a suitable operator \(\rho\) for \(\mathcal{DL}\)
13: while \(\mathcal{D} \neq \emptyset\) do
14: \(C \leftarrow \text{pick}(\mathcal{D})\) % pick a concept \(C\) to be specialised
15: \(\mathcal{D} \leftarrow \mathcal{D}\setminus\{C\}\) % remove \(C\) from the concepts to be specialised
16: \(\mathcal{C} \leftarrow \mathcal{C} \cup \{C\}\) % add \(C\) to the final set
17: \(\rho_C \leftarrow \text{specialise}(C, \rho, \Sigma, \ell_{max})\) % specialise \(C\) using \(\rho\)
18: \(\mathcal{D}_C \leftarrow \{D \in \text{urc}(\rho_C) \mid \exists D' \in \mathcal{C} \cup \mathcal{D} : D' \equiv D\}\) % discard variations
19: \(\mathcal{D} \leftarrow \mathcal{D} \cup \{D \in \mathcal{D}_C \mid p(D, \mathcal{O}) \geq p_{min}\}\) % add suitable specialisations
20: end while
21: return \(\mathcal{C}\)

---

**Example 7.4.** Consider the ontology \(\mathcal{O}\) used in Example 3.1.

\[
\mathcal{O} := \{\text{Man} \lor \neg \text{Woman}, \text{hasParent} \lor \text{hasChild}, \text{Man}(\text{Arthur}), \text{Man}(\text{Chris}), \text{Man}(\text{James}), \text{Woman}(\text{Penelope}), \text{Woman}(\text{Victoria}), \text{Woman}(\text{Charlotte}), \text{Woman}(\text{Margaret}), \text{hasParent}(\text{Charlotte}, \text{James}), \text{hasParent}(\text{Charlotte}, \text{Victoria}), \text{hasParent}(\text{Victoria}, \text{Chris}), \text{hasParent}(\text{Victoria}, \text{Penelope}), \text{hasParent}(\text{Arthur}, \text{Penelope}), \text{hasParent}(\text{Arthur}, \text{Chris})\}.
\]

specialise concepts only if they have \(\geq \ell_{max}\) instances!
Slava implements: DL Miner

![Diagram of DL Miner architecture](image)

**Figure 4.1: Architecture of DL-Miner**

- **Hypothesis Sorter**, given the quality function $qf(\cdot)$, orders hypotheses $H$ according to the binary relation $O$. The result is the ranking function $rf(H)$ that returns the quality rank of a hypothesis $H$.

- The output of DL-Miner is a set $H$ of hypotheses, quality function $qf(\cdot)$, and ranking function $rf(\cdot)$. Domain experts and ontology engineers are supposed to navigate through the hypotheses using the quality and ranking functions. Thus, all hypotheses can be methodically examined. Clearly, it is possible to select only the best hypotheses if necessary. As the reader will find in the following, hypotheses of DL-Miner can, in fact, be used for various purposes and in different scenarios.

In the following, we clarify the parameters and unfold the functionality of each block. Hypothesis Evaluator is covered in Chapter 5, where we define quality measures that can be used in $Q$, and Chapter 6, where we develop techniques to compute those measures. Hypothesis Constructor is explained in Chapter 7 where we show how to construct suitable concepts $C(\cdot)$ from a language $L$ and generate hypotheses $H$ from $C(\cdot)$. Ontology Cleaner and Hypothesis Sorter are both covered in Chapter 8 where we also integrate all techniques in DL-Miner. Finally, we empirically evaluate DL-Miner in Chapter 9.

The symbol "\(\cdot\)" stands for the argument of the function if irrelevant.

When $O$ and $Q$ are clear from the context, we denote the binary relation $O, Q$ by $\cdot$.  

---

**Figure 8.3: Architecture of DL-Miner with subroutines**

Another property of Algorithm 11 is that it always terminates. This ensures that the algorithm returns an output (even though it may take long) for any legal input parameters, i.e. satisfying the respective constraints of Algorithm 11. The properties of correctness, completeness, and termination of Algorithm 11 follow from the same properties of its subroutines, see Theorem 8.1.

**Theorem 8.1 (Correctness, completeness, termination).**

Let $O, L := (DL, `\text{max}', p_{\text{min}}, G, R, n)$ be legal parameters of DL-Miner. Let (i) – (iii) be the following conditions for a hypothesis $H$:

(i) $H$ conforms to $L$;
(ii) $H$ is in NNF;
(iii) $e_H \in \Sigma$.

Then, all following properties hold for DL-Miner:

- it terminates;
- it is correct: it returns a set $H$ of hypotheses such that $H = H$ implies $H$ satisfies (i) – (iii);
4.2. DESIGNING DL-Miner

The output of DL-Miner is a set $H$ of hypotheses, quality function $qf(\cdot)$, and ranking function $rf(\cdot)$. Domain experts and ontology engineers are supposed to navigate through the hypotheses using the quality and ranking functions. Thus, all hypotheses can be methodically examined. Clearly, it is possible to select only best hypotheses if necessary. As the reader will find in the following, hypotheses of DL-Miner can, in fact, be used for various purposes and in different scenarios.

In the following, we clarify the parameters and unfold the functionality of each block. Hypothesis Evaluator is covered in Chapter 5, where we define quality measures that can be used in $Q$, and Chapter 6, where we develop techniques to compute those measures. Hypothesis Constructor is explained in Chapter 7 where we show how to construct suitable concepts $C(\text{roles } R)$ given a language bias $L$ and generate hypotheses $H$ from $C(\text{R})$. Ontology Cleaner and Hypothesis Sorter are both covered in Chapter 8 where we also integrate all techniques in DL-Miner. Finally, we empirically evaluate DL-Miner in Chapter 9.

3 The symbol "\cdot" stands for the argument of the function if it is irrelevant.
4 When $O$ and $Q$ are clear from the context, we denote the binary relation $O, Q$ by $\cdot$.

The symbol "\cdot" stands for the argument of the function if it is irrelevant.

4 When $O$ and $Q$ are clear from the context, we denote the binary relation $O, Q$ by $\cdot$.
DL Miner: Hypothesis Evaluator

- Relatively straightforward for axiom measures
  - hard test case for instance retrieval

- Hard for set-of-axiom measures (fitness & braveness)
  - due to $d\text{Len}(A, O) = \min\{\ell(A') \mid A' \cup O \equiv A \cup O\}$
  - DL Miner implements an approximation that
    - identifies redundant assertions in ABox
      $d\text{Len}^*(A, O) = \ell(A) - \ell(\text{Redundt}(A, O))$
    - does consider 1-step interactions between individuals
    - ignores ‘longer’ interactions
    - underestimates fitness, overestimates braveness
  - great test case for incremental reasoning: Pellet!
DL Miner: Hypothesis Sorter

• Last step in DL Miner’s workflow
• Easy:
  – throw away all hypotheses that are dominated by another one
  – i.e., compute the Pareto front wrt the measures provided
DL Miner: Example

Given a Kinship Ontology,¹ it mines 536 Hs with confidence above 0.9, e.g.

\[
\begin{align*}
\text{Woman} \sqcap \exists \text{hasChild}.\top & \sqsubseteq \text{Mother} \\
\text{Man} \sqcap \exists \text{hasChild}.\top & \sqsubseteq \text{Father} \\
\exists \text{hasChild}.\top & \sqsubseteq \exists \text{marriedTo}.\top \\
\exists \text{marriedTo}.\top & \sqsubseteq \exists \text{hasChild}.\top \\
\exists \text{marriedTo}.\text{Woman} & \sqsubseteq \text{Man} \\
\exists \text{marriedTo}.\text{Mother} & \sqsubseteq \text{Father} \\
\text{Father} & \sqsubseteq \exists \text{marriedTo}.(\exists \text{hasChild}.\top) \\
\text{Mother} & \sqsubseteq \exists \text{marriedTo}.(\exists \text{hasChild}.\top) \\
\exists \text{hasChild}.\top & \sqsubseteq \text{Mother} \sqcup \text{Father} \\
\exists \text{hasChild}.\top & \sqsubseteq \text{Man} \sqcup \text{Woman} \\
\exists \text{hasChild}.\top & \sqsubseteq \text{Father} \sqcup \text{Woman}
\end{align*}
\]

1. adapted from UCI Machine Learning Repository
DL Miner: Example

Given a Kinship Ontology, it mines 536 Hs with confidence above 0.9, e.g.

\[
\begin{align*}
\text{Woman} \sqcap \exists \text{hasChild}. \top & \sqsubseteq \text{Mother} \\
\text{Man} \sqcap \exists \text{hasChild}. \top & \sqsubseteq \text{Father} \\
\exists \text{hasChild}. \top & \sqsubseteq \text{marriedTo}. \top \\
\exists \text{marriedTo}. \top & \sqsubseteq \text{Father} \\
\text{Father} & \sqsubseteq \exists \text{marriedTo}. (\exists \text{hasChild}. \top) \\
\text{Mother} & \sqsubseteq \exists \text{marriedTo}. (\exists \text{hasChild}. \top) \\
\exists \text{hasChild}. \top & \sqsubseteq \text{Mother} \sqcup \text{Father} \\
\exists \text{hasChild}. \top & \sqsubseteq \text{Man} \sqcup \text{Woman} \\
\exists \text{hasChild}. \top & \sqsubseteq \text{Father} \sqcup \text{Woman}
\end{align*}
\]

1. adapted from UCI Machine Learning Repository
Still: many open questions

- If we can compute measure, how feasible is this?
- If “feasible”,
  - do these measures correlate?
  - how independent are they?
- For which DLs & inputs can we create & evaluate hypotheses?
- Which measures indicate *interesting* hypothesis?
- What is the shape of *interesting* hypothesis?
  - are longer/bigger hypotheses better?
- What do we do with them?
  - how do we guide users through these?
Design, run, analyse experiments
Design, run, analyse experiments

- A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     - All BioPortal ontologies with >= 100 individuals and
       >= 100 RAs 21 ontologies
Design, run, analyse experiments

- A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     - All BioPortal ontologies with >= 100 individuals and
       >= 100 RAs 21 ontologies

- Settings for hypothesis parameters:
  - $\mathcal{L}$ is $SHI$
    - RIAs with inverse, composition
  - minsupport = 10
  - max concept length in GCIs = 4
Design, run, analyse experiments

• A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     • All BioPortal ontologies with >= 100 individuals and >= 100 RAs 21 ontologies

• Settings for hypothesis parameters:
  – \( \mathcal{L} \) is \( SHI \)
    – RIAs with inverse, composition
  – minsupport = 10
  – max concept length in GCIs = 4

• generate & evaluate up to 500 hypotheses per ontology
Design, run, analyse experiments

- What kind of axioms do people write?
  - re. readability of hypotheses:
  - what kind of axioms should we roughly aim for?

Use of DL constructors in Bioportal - Taxonomies

<table>
<thead>
<tr>
<th>DL constructor</th>
<th>C</th>
<th>$\exists R.C$</th>
<th>$C \cap D$</th>
<th>$\forall R.C$</th>
<th>$C \cup D$</th>
<th>$\neg C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axioms, %</td>
<td>99.73</td>
<td>67.82</td>
<td>1.15</td>
<td>0.46</td>
<td>0.09</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Length & role depth of axioms in Bioportal - Taxonomies

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>mode</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>2.63</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>depth</td>
<td>0.69</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Design, run, analyse experiments

- What kind of axioms do people write?
  - re. readability of hypotheses:
  - what kind of axioms should we roughly aim for?

Use of DL constructors in Bioportal - Taxonomies

<table>
<thead>
<tr>
<th>DL constructor</th>
<th>Axioms, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>75%</td>
</tr>
<tr>
<td>¬D</td>
<td>25%</td>
</tr>
<tr>
<td>¬C</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Length & role depth of axioms in Bioportal - Taxonomies

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>mode</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>2.63</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>depth</td>
<td>0.69</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Restricting length of concepts in axioms to 4 (axioms to 8) is fine!
Design, run, analyse experiments

How do the measures correlate?

Figure 9.1: Mutual correlations of quality measures for handpicked (a) and principled (b) corpus: positive correlations are in blue, negative correlations are in red, crosses mark statistically insignificant correlations (significance level 0.05)
Design, run, analyse experiments

How do the measures correlate?

Figure 9.1: Mutual correlations of quality measures for handpicked (a) and principled (b) corpus: positive correlations are in blue, negative correlations are in red, crosses mark statistically insignificant correlations (significance level 0.05)
How do the measures correlate?

(a) Handpicked corpus

(b) Principled corpus
Design, run, analyse experiments

How do the measures correlate?

Figure 9.1: Mutual correlations of quality measures for handpicked (a) and principled (b) corpus: positive correlations are in blue, negative correlations are in red, crosses mark statistically insignificant correlations (significance level 0.05)
Design, run, analyse experiments

How feasible is hypothesis mining?

According to Figure 9.2a, consistency is the most expensive measure which is rather unexpected. Please recall that consistency tests whether the union of a hypothesis and the ontology is consistent. Hence, it can be costly if the ontology is large and/or expressive. Indeed, consistency is considerably more costly for the principled corpus than for the handpicked one which is likely to be a consequence of higher expressivity of the former. The higher cost of consistency for the principled corpus decreases the relative contributions of other measures for this corpus, i.e. fitness, logical strength, etc.

The relatively high computational cost of logical strength is explained by the fact that its performance is measured by comparing a given hypothesis to all others (in the worst case). Hence, it grows with the number of hypotheses to be evaluated. Therefore, it should be compared with other measures cautiously.

As Figure 9.2a shows, considering negation in the statistical axiom measures is relatively expensive. Therefore, given the strong correlation between the basic and main measures, see Figure 9.1, it is sensible to replace the main measures with their basic counterparts in certain cases, particularly if the ontology does not contain negative information in the ABox. On the other hand, if other expensive measures need to be computed, the relative cost of computing all axiom measures is not so big.

The abbreviations in Figure 9.2b stand for the respective running times as follows: OC – ontology parsing and classification; HC – hypotheses construction; Prep. Ent. Checks – preparatory entailment checks; Hypothesis Evaluation – hypothesis evaluation.
Design, run, analyse experiments

How feasible is hypothesis mining?

Works fine for classifiable ontologies. Incremental Reasoning in Pellet works very well for ABoxes.
According to Figure 9.2a, consistency is the most expensive measure which is rather unexpected. Please recall that consistency tests whether the union of a hypothesis and the ontology is consistent. Hence, it can be costly if the ontology is large and/or expressive. Indeed, consistency is considerably more costly for the principled corpus than for the handpicked one which is likely to be a consequence of higher expressivity of the former. The higher cost of consistency for the principled corpus decreases the relative contributions of other measures for this corpus, i.e. fitness, logical strength, etc.

The relatively high computational cost of logical strength is explained by the fact that its performance is measured by comparing a given hypothesis to all others (in the worst case). Hence, it grows with the number of hypotheses to be evaluated. Therefore, it should be compared with other measures cautiously.

As Figure 9.2a shows, considering negation in the statistical axiom measures is relatively expensive. Therefore, given the strong correlation between the basic and main measures, see Figure 9.1, it is sensible to replace the main measures with their basic counterparts in certain cases, particularly if the ontology does not contain negative information in the ABox. On the other hand, if other expensive measures need to be computed, the relative cost of computing all axiom measures is not so big.

The abbreviations in Figure 9.2b stand for the respective running times as follows: OC – ontology parsing and classification; HC – hypotheses construction; FITN – fitness; BRAV – brave; CONS – consistency; INFOR – information; STREN – strength; REDUN – redundancy; DISSIM – dissimilarity; COMPL – completeness.
Design, run, analyse experiments

How costly are the different measures?

Consistency is the most costly measure
But - what about the semantic mining?

TBox

ABox

DL Miner

Hypotheses

axiom(s) m1,m2,m3,
So, what have we got? (new version)

✓ Loads of measures to capture aspects of hypotheses
  – mostly independent
  – some superfluous on positive data (unsurprisingly)

✓ Hypothesis generation & evaluation is feasible
  – provided our ontology is classifiable
  – provided our search space isn’t too massive
    • …focus!

• Which measures indicate *interesting* hypothesis?
• What is the shape for *interesting* hypothesis?
  – are longer/bigger hypotheses better?
• What do we do with them?
  – how do we guide users through these?
Design, run, analyse survey

Can we learn hypotheses are
  • useful/interesting?
  …and how does this correlate with measures?

TBox
SROIQ
$|Ci| \leq 4$
sig = 522

ABox
169K CAs
405K RAs

$SHI$

S1: 60 Hypos unfocused
S2: 60 Hypos focused

DL Miner

axiom(s)
m1, m2, m3,
Design, run, analyse survey

Can we learn hypotheses are
  • usefull/interesting?
…and how does this correlate with measures?

TBox
  SROIQ
  sig = 522

ABox
  169K CAs
  405K RAs

\( SHI \) \(|C_i| \leq 4 \)

S1: 60 Hypos unfocused
S2: 60 Hypos focused

30 high-confidence
30 low-confidence
Design, run, analyse survey

Can we learn hypotheses are
  • useful/interesting?
  …and how does this correlate with measures?

\[ SHI \leq 4 \]

S1: 60 Hypos unfocused
S2: 60 Hypos focused

Valid? Interesting?

TBox

\[ SROIQ \]

sig = 522

ABox

169K CAs
405K RAs

DL Miner

axiom(s)
m1, m2, m3,
How good/valid are the mined hypotheses?

<table>
<thead>
<tr>
<th>Survey 1</th>
<th>Validity</th>
<th>Interestingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>(unfocused)</td>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td></td>
<td>Wrong</td>
<td>6 11 30 - -</td>
</tr>
<tr>
<td></td>
<td>Don’t know</td>
<td>- 1 - 2 4</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>- - - 6 -</td>
</tr>
</tbody>
</table>
## Design, run, analyse survey

### How good/valid are the mined hypotheses?

<table>
<thead>
<tr>
<th>Survey 1 (unfocused)</th>
<th>Validity</th>
<th>Interestingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong</td>
<td>6</td>
<td>11 30</td>
</tr>
<tr>
<td>Don’t know</td>
<td>-</td>
<td>1 - 2 4</td>
</tr>
<tr>
<td>Correct</td>
<td>-</td>
<td>- 6 -</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survey 2 (focused)</th>
<th>Validity</th>
<th>Interestingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong</td>
<td>1</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Don’t know</td>
<td>-</td>
<td>- - 49</td>
</tr>
<tr>
<td>Correct</td>
<td>-</td>
<td>- - 4</td>
</tr>
</tbody>
</table>

As Table 9.10 shows, Survey 1 contains 47 hypotheses deemed to be wrong, 7 hypotheses unknown, and 6 hypotheses correct. The majority of wrong hypotheses are of average interestingness (marked by 2) and the rest of wrong hypotheses are less interesting (marked by 0 or 1). As the domain expert points out in her feedback, wrong hypotheses which are marked by the interestingness of 2 indicate data bias, i.e. those are incorrect but strongly supported by the data. According to the results, unknown and correct hypotheses appear to be much more interesting than wrong ones: all of them, except one, have high values of interestingness (marked by 3 and 4). Amongst those, unknown hypotheses are marked to be the most interesting and, according to the expert’s response, require further analysis. Overall, 12 out of 60 hypotheses (20%) are found to be interesting.

The results of Survey 2 are much different from the results of Survey 1, see Table 9.10. While most hypotheses in Survey 1 are deemed to be wrong, most hypotheses (49 out of 60) in Survey 2 are marked as unknown. Another noticeable difference is that all hypotheses, except two, in Survey 2 are marked by the highest value of interestingness, i.e. 58 out of 60 hypotheses ($\approx 96.7\%$), including wrong ones, are very interesting in expert’s opinion. Moreover, the expert informed us in her response that one of the wrong hypotheses, besides indicating data bias, revealed an error in the ontology. Thus, if focus terms are specified by the domain expert, the resulting focused hypotheses appear to be significantly more interesting than unfocused ones. This is not surprising because, by providing focus terms, the expert expresses her interest in exploring hypotheses about those terms. In addition, the expert is likely to inquire into the domain area which she knows less about. As a result, the majority of focused hypotheses are deemed to be...
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

![Graphs showing correlations between hypothesis quality measures and expert's judgements.](image)

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of "specific" hypotheses. Such hypotheses are likely to have a higher lift than "general" hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

Figure 9.8: Correlations (in descending order) between hypothesis quality measures and expert's judgements (4 measures are not shown for Survey 2 because their deviations equal zero and correlation coefficients cannot be calculated)

likely to be “general”, i.e. reflecting known, easily seen patterns of the data. Those hypotheses are not as surprising as “specific” ones which, on the contrary, are likely to reflect uncommon, hardly seen patterns of the data.

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of “specific” hypotheses. Such hypotheses are likely to have a higher lift than “general” hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

Figure 9.8: Correlations (in descending order) between hypothesis quality measures and expert’s judgements (4 measures are not shown for Survey 2 because their deviations equal zero and correlation coefficients cannot be calculated)

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of “specific” hypotheses. Such hypotheses are likely to have a higher lift than “general” hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to...
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

![Correlation coefficients graph]

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of “specific” hypotheses. Such hypotheses are likely to have a higher lift than “general” hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to...
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

Figure 9.8: Correlations (in descending order) between hypothesis quality measures and expert’s judgements (4 measures are not shown for Survey 2 because their deviations equal zero and correlation coefficients cannot be calculated).

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of “specific” hypotheses. Such hypotheses are likely to have a higher lift than “general” hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to...
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

Figure 9.8: Correlations (in descending order) between hypothesis quality measures and expert's judgements (4 measures are not shown for Survey 2 because their deviations equal zero and correlation coefficients cannot be calculated).

The results in Figure 9.8c are similar to the results in Figure 9.8a: the correlations for validity look almost equally distributed. The main difference is that lift turns from a non-indicator in Survey 1 to a strong positive indicator in Survey 2. One possible reason is that Survey 2 consists of "specific" hypotheses. Such hypotheses are likely to have a higher lift than "general" hypotheses in Survey 1 (since the denominator decreases faster than the numerator, see Definition 5.16).

In comparison to Figure 9.8b for Survey 1, Figure 9.8d for Survey 2 shows considerably stronger correlations for interestingness. This may be caused by the much higher fraction of interesting hypotheses in Survey 2 in comparison to...
What we learned: 3 kinds of hypotheses!

1. An interesting hypothesis can give *new insights into domain*

TBox → DL Miner → ABox

Hypotheses:
- axiom(s) m1, m2, m3,
- high confidence/lift/…
- low assumptions/braveness
What we learned: 3 kinds of hypotheses!

1. An interesting hypothesis can give new insights into domain
What we learned: 3 kinds of hypotheses!

2. An interesting hypothesis can reveal *axioms missing from TBox*!!
What we learned: 3 kinds of hypotheses!

2. An interesting hypothesis can reveal *axioms missing from TBox*

*TBox completion*  
*ontology learning from data*
What we learned: 3 kinds of hypotheses!

3. An interesting hypothesis can reveal *bias & errors in the ontology*

- **TBox**
- **ABox**

**DL Miner**

**Hypotheses**

- **axiom(s):** m1,m2,m3,
- high confidence/lift/…
- low assumptions/braveness
What we learned: 3 kinds of hypotheses!

3. An interesting hypothesis can reveal bias & errors in the ontology.

Semantic Data Analysis
3 kinds of hypotheses - can we predict?

No - they look alike

TBox

ABox

DL Miner

Hypotheses

axiom(s) m1, m2, m3,

high confidence/lift/…
low assumptions/braveness
3 kinds of hypotheses - can we predict?

No - they look alike
Perhaps - with different ABoxes/other sources
Summary & Outlook

- Mining rich axioms from ontologies is possible
  - gives us more than we thought
  - expressive axioms are better!
- Fine test case for incremental/ABox reasoning
- More surveys
  - to better understand relevance of metrics
  - but we’ve got the shape now
- Redundancy in general is tricky & costly
  - stripping superfluous parts from concepts, (sets of) axioms
- We need even better refinement operators:
  - for more expressive DLs
  - redundancy-free
  - ontology-aware
Subjective ontology-based problems

- are great fun
  - design of experiments & surveys
  - but also rather complex: sooo many design choices
- specifying & implementing good parameters is tricky
  - metrics make “ontology mining” subjective
  - requires understanding of logic & reasoners & …
- are plentiful/numerous
  - abduction
  - similarity
  - good explanations/proofs for entailments justifications
  - good counter-models for non-entailments
  - good repair of inconsistent/incoherent ontologies
  - …
Special Thanks to Slava Sazonau
Description Logics (DLs) have a long tradition in computer science and knowledge representation, being designed so that domain knowledge can be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.