Outline

Introduction and Motivation

Generating Verbalized Explanations

Experimental Study

Conclusion & Discussion
Motivation

Knowledge representation essential for

- information management systems,
- intelligent systems,
- cognitive systems,
- Semantic Web,
- ...
Logic-Based Knowledge Representation

- Flexible representation of data and schema
- Well-defined semantics

→ (Automated) deduction support:
  - Derivation of logical consequences, i.e., implicit knowledge is made explicit
- Often via ontologies formalized in Description Logics (DLs)
- DLs underpin the W3C standardized Web Ontology Language OWL
Explanations of Inferences

- Strength of logic-based methods: Logical inferences can be explained

  - **Justifications**: A minimal set of axioms that support a conclusion
    - Problem: Users unfamiliar with logic and inference calculi

  - **Explanations**: *Step-wise* explanations for inferences as *natural language*
    - Problem: Explanations can become very long
      - Techniques to make explanations more *concise*
      - Empirical evidence that verbalizations are (still) *understandable*

- Related work:
  - [Horridge et al., 2011]: Empirical study on complexity of justifications
  - [Nguyen et al., 2012]: Empirical study on understandability of individual inference steps
  - [Kazakov & Klinov, 2014]: Consequence-based reasoning with tracing
### Example

**Axioms**

\[
\text{EsophagealPathology} \equiv (\text{PathologicalCondition} \sqcap \exists \text{LocativeAttribute. Esophagus})
\]

\[
\text{Esophagus} \sqsubseteq \text{GastrointestinalTractBodyPart}
\]

\[
\text{DigestiveSystemPathology} \equiv (\text{PathologicalCondition} \sqcap \exists \text{LocativeAttribute. GastrointestinalTractBodyPart})
\]

\[
\models
\]

**Subsumption**

\[
\text{EsophagealPathology} \sqsubseteq \text{DigestiveSystemPathology}
\]

**Verbalization**

An *esophageal pathology* is *defined* as a pathological condition that is located in the esophagus. The *esophagus* is a part of the *gastrointestinal tract*, thus an *esophageal pathology* is located in a part of the gastrointestinal tract.

Furthermore, since an *esophageal pathology* is a pathological condition, an *esophageal pathology* is a pathological condition that is located in a part of the gastrointestinal tract.

A *digestive system pathology* is *defined* as a pathological condition that is located in a part of the gastrointestinal tract. Thus, an *esophageal pathology* is a digestive system pathology.
Considered Ontology Language

- Description Logic $\mathcal{EL}$ plus common features underlying the OWL2 EL profile:
  - unsatisfiable concept
  - role inclusion
  - role composition
  - domain axioms
  - disjointness axioms
- OWL2 EL covers a number of large and practically relevant ontologies (e.g. SNOMED CT, GO)
- Considered inferences: Subsumption between concepts
- Challenge: Compute concise and understandable explanations
Preliminaries – Considered $\mathcal{EL}$ Fragment

### Names
- Class names: $N_C = \{A, B, C, \ldots\}$
- Individual names: $N_I = \{a, b, c, \ldots\}$
- Role names: $N_R = \{r, s, t, \ldots\}$

### Interpretation
- $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where
- $\Delta^\mathcal{I}$ is a nonempty set of individuals
- $\cdot^\mathcal{I} : N_C \rightarrow \mathcal{P}(\Delta^\mathcal{I})$, $N_I \rightarrow \Delta^\mathcal{I}$, $N_R \rightarrow \mathcal{P}(\Delta^\mathcal{I} \times \Delta^\mathcal{I})$

### $\mathcal{EL}$ Class Expressions & Axioms

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomic class</td>
<td>$C$</td>
</tr>
<tr>
<td>Universal concept</td>
<td>$\top$</td>
</tr>
<tr>
<td>Intersection</td>
<td>$C_1 \cap C_2$</td>
</tr>
<tr>
<td>Existential restriction</td>
<td>$\exists r.C_1$</td>
</tr>
<tr>
<td>Subsumption axiom</td>
<td>$C_1 \subseteq C_2$</td>
</tr>
</tbody>
</table>

### Further Elements from $\mathcal{EL}^{++}$

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsatisfiable concept</td>
<td>$\bot$</td>
</tr>
<tr>
<td>Nominal</td>
<td>${a}$</td>
</tr>
<tr>
<td>Role composition</td>
<td>$r_1 \circ \ldots \circ r_k \subseteq s$</td>
</tr>
<tr>
<td>Domain axiom</td>
<td>$\text{dom}(r, C_1)$</td>
</tr>
<tr>
<td>Disjointness axiom</td>
<td>$\text{disj}(C_1, C_2)$</td>
</tr>
</tbody>
</table>

... where $C_1$ and $C_2$ are class expressions
Verbalization Pipeline for a Given Subsumption

1. Compute justifications [Kalyanpur, 2006; Horridge, 2011]
Verbalization Pipeline for a Given Subsumption

1. Compute justifications [Kalyanpur, 2006; Horridge, 2011]
2. Construct rule-based proof
   ▶ Using custom ruleset (includes/modifies rules from ELK and Nguyen et al.)
Verbalization Pipeline for a Given Subsumption

1. Compute justifications [Kalyanpur, 2006; Horridge, 2011]
2. Construct rule-based proof
   - Using custom ruleset (includes/modifies rules from ELK and Nguyen et al.)

For the example:

\[
\begin{align*}
R_{\equiv} & \quad EP \equiv PC \sqcap \exists loc.E \\
R_{\equiv} & \quad EP \sqsubseteq PC \\
R_{15} / R_5 & \quad EP \sqsubseteq PC \sqcap \exists loc.E \\
R_{\equiv} & \quad EP \sqsubseteq PC \sqcap \exists loc.GTP \\
R_{\equiv} & \quad EP \sqsubseteq DSP \\
\end{align*}
\]

(EP: EsophagealPathology, E: Esophagus, PC: PathologicalCondition, GTP: GastrointestinalTractBodyPart, DSP: DigestiveSystemPathology, loc: LocativeAttribute)
Verbalization Pipeline for a Given Subsumption

1. Compute justifications [Kalyanpur, 2006; Horridge, 2011]
2. Construct rule-based proof
   ▶ Using custom ruleset (includes/modifies rules from ELK and Nguyen et al.)

For the example:

\[
\begin{align*}
R_{\equiv} & \quad EP \equiv PC \sqcap \exists \text{loc.}E \\
R_{\equiv} & \quad EP \sqsubseteq PC \\
R_{\equiv} & \quad EP \equiv PC \sqcap \exists \text{loc.}GTP \\
R_{\equiv} & \quad EP \sqsubseteq PC \sqcap \exists \text{loc.}GTP
\end{align*}
\]

\[
\begin{align*}
R_{\equiv} & \quad EP \equiv PC \sqcap \exists \text{loc.}E \\
R_{\equiv} & \quad EP \sqsubseteq \exists \text{loc.}E \\
R_{\equiv} & \quad EP \equiv \exists \text{loc.}GTP \\
R_{\equiv} & \quad EP \sqsubseteq \exists \text{loc.}GTP \\
R_{\equiv} & \quad EP \equiv DSP \\
R_{\equiv} & \quad EP \sqsubseteq DSP
\end{align*}
\]

(EP: EsophagealPathology, E: Esophagus, PC: PathologicalCondition, GTP: GastrointestinalTractBodyPart, DSP: DigestiveSystemPathology, loc: LocativeAttribute)

3. Convert prooftree to text using patterns
Verbalizing Inference Rules

- Post-order traversal of proof tree
- Rules specify order in which premises are verbalized
- Patterns do not repeat premises verbalized as conclusion of previous steps

Example

\[
\text{R}_\sqsubseteq/R_{12} \quad \frac{(1) \ C_1 \sqsubseteq C_2 \quad (2) \ C_2 \sqsubseteq C_3}{C_1 \sqsubseteq C_3}
\]

(1)&(2) *Since* \( v(C_1 \sqsubseteq C_2) \) *and* \( v(C_2 \sqsubseteq C_3) \), it follows that \( v(C_1 \sqsubseteq C_3) \).

(1) \( v(C_1 \sqsubseteq C_2) \), *therefore being* \( v(C_3) \).

(2) *Given that* \( v(C_2 \sqsubseteq C_3) \), \( v(C_1 \sqsubseteq C_3) \).

---

Thus, we have established that \( v(C_1 \sqsubseteq C_3) \).
Verbalizing Class Expressions and Axioms

Recursively apply verbalization patterns to class expressions. The basic patterns are:

<table>
<thead>
<tr>
<th>Class Expression</th>
<th>Pattern</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>( a[n] A )</td>
<td>a mollusc</td>
</tr>
<tr>
<td>( A \cap B )</td>
<td>( v(A) ) that is ( v(B) ) [...and...]</td>
<td>a mollusc that is a cephalopod</td>
</tr>
<tr>
<td>( \exists r.A )</td>
<td>something that ( r ) ( v(A) ) [... and...]</td>
<td>something that has a shell</td>
</tr>
<tr>
<td>( A \subseteq B )</td>
<td>( v(A) ) is ( v(B) )</td>
<td>a clam is a mollusc</td>
</tr>
<tr>
<td>( A \equiv B )</td>
<td>According to its definiti-</td>
<td>According to its definiti-</td>
</tr>
<tr>
<td></td>
<td>on, ( v(A) ) is ( v(B) )</td>
<td>on, a typical mollusc is</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[...]</td>
</tr>
</tbody>
</table>
Optimization: Aggregating Class Expressions

- For intersections of class names:
  - \( \text{old	extunderscore lady} \sqsubseteq \text{elderly} \sqcap \text{female} \sqcap \text{person} \)
  - \( \leadsto \) “An old lady is an elderly that is a female and a person.”
  - \( \leadsto \) “An old lady is an elderly female person.”

- For existential restrictions on the same role:
  - \( \text{Saline} \sqsubseteq \exists \text{hasDissolvedWithin}. \text{Chlorine} \sqcap \exists \text{hasDissolvedWithin}. \text{Sodium} \)
  - \( \leadsto \) “Saline is something that has dissolved within chlorine and something that has dissolved within sodium”
  - \( \leadsto \) “Saline is something that has dissolved within chlorine and sodium”

- Furthermore, unnamed concepts are named, if possible (using domain axioms):
  - “something that owns a dog”
  - \( \leadsto \) “a person that owns a dog”
Further Techniques & Heuristics to Promote Conciseness

- Inference rules that represent logical “shortcuts”, e.g.:
Further Techniques & Heuristics to Promote Conciseness

▶ Inference rules that represent logical “shortcuts”, e.g.:

\[
\begin{align*}
R \equiv & \frac{C_1 \sqsubseteq C_2 \quad C_2 \equiv C_3}{C_1 \sqsubseteq C_3} \quad \text{combines} \quad R_1 \frac{C_1 \equiv C_2}{C_2 \sqsubseteq C_2} + R \equiv /R_{12} \frac{C_1 \sqsubseteq C_2 \quad C_2 \sqsubseteq C_3}{C_1 \sqsubseteq C_3}
\end{align*}
\]
Further Techniques & Heuristics to Promote Conciseness

▶ Inference rules that represent logical “shortcuts”, e.g.:

\[ R \equiv \frac{c_1 \sqsubseteq c_2 \quad c_2 \equiv c_3}{c_1 \sqsubseteq c_3} \quad \text{combines} \quad R_1 \frac{c_1 \equiv c_2}{c_2 \sqsubseteq c_2} + \quad R \equiv /R_{12} \frac{c_1 \sqsubseteq c_2 \quad c_2 \sqsubseteq c_3}{c_1 \sqsubseteq c_3} \]

\[ R_r^{+}/R_5 \frac{c_1 \sqsubseteq c_2 \ldots c_1 \sqsubseteq c_{n+1}}{c_1 \sqsubseteq c_2 \sqcap \ldots \sqcap c_{n+1}} \quad \text{(n-ary)} \]
Further Techniques & Heuristics to Promote Conciseness

- Inference rules that represent logical “shortcuts”, e.g.:

\[
\begin{align*}
R \equiv & \begin{array}{c}
C_1 \sqsubseteq C_2 \quad C_2 \equiv C_3 \\
C_1 \sqsubseteq C_3
\end{array} \\
\text{combines} & \quad R_1 \begin{array}{c}
C_1 \equiv C_2 \\
C_2 \sqsubseteq C_2
\end{array} + \quad R \equiv / R_{12} \begin{array}{c}
C_1 \sqsubseteq C_2 \\
C_2 \sqsubseteq C_3
\end{array}
\end{align*}
\]

\[
R_{1+}/R_5 \begin{array}{c}
C_1 \sqsubseteq C_2 \quad \ldots \quad C_1 \sqsubseteq C_{n+1} \\
C_1 \sqsubseteq C_2 \sqcap \ldots \sqcap C_{n+1}
\end{array} \quad (n\text{-ary})
\]

- Skipping of “trivial” inference steps in the verbalization
Further Techniques & Heuristics to Promote Conciseness

- Inference rules that represent logical “shortcuts”, e.g.:

  \[ R \equiv C_1 \sqsubseteq C_2, C_2 \equiv C_3 \quad \text{combines} \quad R_1 \frac{C_1 \equiv C_2}{C_2 \sqsubseteq C_2} \quad + \quad R \equiv \frac{R_1 \sqcup / R_{12}}{C_1 \sqsubseteq C_2} \frac{C_2 \sqsubseteq C_3}{C_1 \sqsubseteq C_3} \]

  \[ R_{\sqcap}^{+} / R_5 \frac{C_1 \sqsubseteq C_2 \ldots C_1 \sqsubseteq C_{n+1}}{C_1 \sqsubseteq C_2 \sqcap \ldots \sqcap C_{n+1}} \quad \text{(n-ary)} \]

- Skipping of “trivial” inference steps in the verbalization

  \[ R_{\sqcap}^{-} / R_2 \frac{C_1 \sqsubseteq C_2 \sqcap \ldots \sqcap C_n}{C_1 \sqsubseteq C_i} \quad 2 \leq i \leq n \]

  (since this would yield, e.g. “A man is a human that is a male.”
  \[ \rightarrow \text{“A man is a human.”} \]
Further Techniques & Heuristics to Promote Conciseness

- Inference rules that represent logical “shortcuts”, e.g.:

\[
R \equiv \begin{array}{c}
\frac{C_1 \sqsubseteq C_2 \quad C_2 \equiv C_3}{C_1 \sqsubseteq C_3}
\end{array}
\text{combines}

\[
R_1 \quad \frac{C_1 \equiv C_2}{C_2 \sqsubseteq C_2} \quad + \quad R_2 / R_{12} \quad \frac{C_1 \sqsubseteq C_2 \quad C_2 \sqsubseteq C_3}{C_1 \sqsubseteq C_3}
\]

\[
R^+ / R_5 \quad \frac{C_1 \sqsubseteq C_2 \quad \ldots \quad C_1 \sqsubseteq C_{n+1}}{C_1 \sqsubseteq C_2 \cap \ldots \cap C_{n+1}} \quad \text{(n-ary)}
\]

- Skipping of “trivial” inference steps in the verbalization

\[
R^- / R_2 \quad \frac{C_1 \sqsubseteq C_2 \cap \ldots \cap C_n}{C_1 \sqsubseteq C_i} \quad 2 \leq i \leq n
\]

(since this would yield, e.g. “A man is a human that is a male.” → “A man is a human.”)

- Verbalization patterns designed to avoid textual repetitions
Experimental Study: Goals

- Measure adequacy of verbalizations (Is text readable and natural?)
- Measure “understandability”
- Compare two modes of verbalization: proof-based explanation (cf. previous slides) vs. verbalized justifications (axioms verbalized one-by-one), e.g.

  A vibrio vaccine is a bacterial vaccine.
  A vibrio cholerae vaccine is a vibrio vaccine.
  Vaxchora uses a live attenuated pathogen.
  Vaxchora is a vibrio cholerae vaccine.
  A bacterial vaccine is a vaccine.
  According to its definition, a live attenuated vaccine is a vaccine that uses a live attenuated pathogen.
  Thus, vaxchora is a live attenuated vaccine.
Experimental Study: Procedure

- Recruitment of “reliable” Amazon Mechanical Turk ‘Masters’ ($n = 24$)
- Participants judged four verbalized justifications and four explanations each (3–4 inference steps, 3–6 premise axioms)
- Four groups of participants (random assignment), each judged
  - 2 correct and 2 faulty explanations (incl. shortcut rules, skipping)
  - 2 correct and 2 faulty verbalized justifications
- Before the experiment: pre-study questionnaire (English language skills, age, familiarity with used domains), instructions using an example verbalization
- After the experiment: concluding questions (sufficient allocated time)
- Real-world ontologies from diverse domains: food, vaccines, material, (bio)medicine, physiology
Experimental Study: Instructions

Based on his knowledge base, Toby determines that the following conclusion should hold:

A **nautiloid** is a **typical mollusc**.

Toby provides the following explanation for his conclusion:

- A **nautiloid** is a **cephalopod**
- A **cephalopod** is a **mollusc**
- A **nautiloid** has a **shell**

According to its definition, a **typical mollusc** is a **mollusc** that has a **shell**.

Thus, a **nautiloid** is a **typical mollusc**.

The statements mentioned by Toby in the underlined part of the explanation are true according to his knowledge base. Toby claims the underlined statements logically entail his conclusion. So, is this the case?

In the example above, Toby’s reasoning is logically correct. That is, it is impossible that the statements in Toby’s knowledge base are true while the conclusion is false.

However, Toby sometimes makes errors when he reasons. The following example shows what an error could look like.

**Conclusion**

A **sediment** is a **solid substance**.

**Explanation**

- **Quartz** is a **mineral**.
- A **mineral** is a **solid substance**.

It follows that a **sediment** is a **solid substance**.

In the example above, Toby’s reasoning contains a logical error because his conclusion can be false while the underlined statements from his knowledge base are true. That is, even if quartz is a mineral and even if a mineral is a solid substance, it is still possible that a sediment is not a solid substance.
Experimental Study: Example Explanation

Based on his knowledge base, Toby determines that the following conclusion should hold:

Vaxchora is a live attenuated vaccine.

Toby provides the following explanation for his conclusion. Assume that each individual statement mentioned by Toby in the underlined part of the explanation is true according to his knowledge base.

Since vaxchora is a vibrio cholerae vaccine, which is a vibrio vaccine, which is a bacterial vaccine, which is a vaccine, vaxchora is a vaccine. Furthermore, since vaxchora uses an oral suspension, vaxchora is a vaccine that uses a live attenuated pathogen.

A live attenuated vaccine is defined as a vaccine that uses a live attenuated pathogen. Thus, vaxchora is a live attenuated vaccine.
## Experimental Study: Example Questions

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Slightly disagree</th>
<th>Neutral</th>
<th>Slightly agree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The presented reasoning is logically correct (i.e. Toby’s conclusion is a logical consequence of the available knowledge).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am confident with my answer as to whether the reasoning in the explanation is correct.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The reasoning presented in the explanation is understandable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The explanation conveys less information than I need to fully understand it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I find that the explanation should be made more concise.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The text of the explanation is well-formed (according to writing conventions).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I find the explanation easy to read.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The explanation or parts of it are ambiguous.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is clear how the statements in the explanation are to be interpreted.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The sentences are arranged such that they fit together well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- 71% accuracy for correct vs. incorrect inferences (same for explanations and justifications, mostly incorrect classified as correct $\leadsto$ not read carefully, 4 participants excluded)

- Results consider logically correct explanations that were accepted as correct
Results: Mean Response

- **Strongly agree**
  - exp
  - jus

- **Strongly disagree**

- ◊ mean values
- boxes upper/lower quartile
- – median
Results: Mean Response

Q1 I am confident with my answer as to whether the reasoning in the explanation is correct.

Q2 The reasoning presented in the explanation is understandable.

Q3 The explanation conveys less information than I need to fully understand it.
Results: Mean Response

Q4 I find that the explanation should be made more concise.

Q5 The text of the explanation is well-formed (according to writing conventions). (trend* of preference for explanations)

Q6 I find the explanation easy to read. (trend* of preference for explanations)

*In both cases, Wilcoxon signed rank test p-value=0.08
Results: Mean Response

Q7 The explanation or part of it are ambiguous.

Q8 It is clear how the statements in the explanation are to be interpreted.

Q9 The sentences are arranged such that they fit together well. (trend* of preference for explanations)

*Wilcoxon signed rank test p-value=0.08
Conclusion & Discussion

- Shortened explanations are (slightly) preferred
- Still relatively few inferences evaluated with few participants
- Used in a DIY assistant (developed with Robert Bosch GmbH)