Semantic Query Extension through Probabilistic Description Logics

José Ochoa Luna¹  Kate Revoredo²  Fabio Cozman¹

Decision Making Lab.
Escola Politécnica - Universidade de São Paulo

Departamento de Informática Aplicada
Unirio

URSW 2010
Outline

1. Introduction
2. Probabilistic Description Logic CR\textit{ALC}
3. Semantic Query Extension with CR\textit{ALC}
4. Preliminary Results
5. Conclusions
6. Future Work
Focus: use of ontologies to improve keyword-based search
Focus: use of ontologies to improve keyword-based search

An ontology can be employed for semantic query extension
Methods

Semantic Query Extension

Identification of semantic concepts contained in user queries
Methods

Semantic Query Extension
Identification of semantic concepts contained in user queries

Probabilistic Ontology
It may not be possible to guarantee that a concept is related to a query → uncertainty → (PDL) CRALC
Idea

To obtain all concept instances that are related to a given word even if that word does not appear with the concept.
Idea

To obtain all concept instances that are related to a given word even if that word does not appear with the concept

\[ P(\text{Concept}|\text{Query}) \]
Probabilistic Description Logic $\mathcal{CRALC}$

- $\mathcal{CRALC}$ is a probabilistic extension of the DL $\mathcal{ALC}$. 
**Probabilistic Description Logic CR\textsc{ALC}**

- CR\textsc{ALC} is a probabilistic extension of the DL \textsc{ALC}.
- The following constructors are available in \textsc{ALC}: conjunction \textit{(C \sqcap D)}, disjunction \textit{C \sqcup D}, negation \textit{(\neg C)}, existential restriction \textit{(\exists r.C)}, and value restriction \textit{(\forall r.C)}.
$P(A|B) = \alpha$

$\forall x \in D: P(A(x)|B(x)) = \alpha$

$P(Professor(Maria)|Researcher(Maria)) = 0.4$
Probabilistic Inclusions and their Semantics

- $P(A|B) = \alpha$
- $\forall x \in \mathcal{D} : P(A(x)|B(x)) = \alpha$
Probabilistic Inclusions and their Semantics

- \( P(A|B) = \alpha \)
- \( \forall x \in D : P(A(x)|B(x)) = \alpha \)
- \( P(\text{Professor(Maria)}|\text{Researcher(Maria)}) = 0.4 \)
Example

\[
P(\text{Animal}) = 0.9, \\
P(\text{Rational}) = 0.6, \\
P(\text{hasChild}) = 0.3, \\
\text{Human} \equiv \text{Animal} \cap \text{Rational}, \\
\text{Beast} \equiv \text{Animal} \cap \neg \text{Rational}, \\
\text{Parent} \equiv \\
\text{Human} \cap \exists \text{hasChild}.\text{Human}, \\
P(\text{Kangaroo}|\text{Beast}) = 0.4, \\
P(\text{Kangaroo}|\neg\text{Beast}) = 0.0, \\
\text{MaternityKangaroo} \equiv \\
\text{Kangaroo} \cap \exists \text{hasChild}.\text{Kangaroo}
\]
Example

\[ P(\text{Animal}) = 0.9, \]
\[ P(\text{Rational}) = 0.6, \]
\[ P(\text{hasChild}) = 0.3, \]
\[ \text{Human} \equiv \text{Animal} \cap \text{Rational}, \]
\[ \text{Beast} \equiv \text{Animal} \cap \neg \text{Rational}, \]
\[ \text{Parent} \equiv \]
\[ \text{Human} \cap \exists \text{hasChild}.\text{Human}, \]
\[ P(\text{Kangaroo} | \text{Beast}) = 0.4, \]
\[ P(\text{Kangaroo} | \neg \text{Beast}) = 0.0, \]
\[ \text{MaternityKangaroo} \equiv \]
\[ \text{Kangaroo} \cap \exists \text{hasChild}.\text{Kangaroo} \]
Example

\[ P(\text{Animal}) = 0.9, \]
\[ P(\text{Rational}) = 0.6, \]
\[ P(\text{hasChild}) = 0.3, \]

\[ \text{Human} \equiv \text{Animal} \cap \text{Rational}, \]
\[ \text{Beast} \equiv \text{Animal} \cap \neg \text{Rational}, \]

\[ \text{Parent} \equiv \]
\[ \text{Human} \cap \exists \text{hasChild.Human}, \]
\[ P(\text{Kangaroo}|\text{Beast}) = 0.4, \]
\[ P(\text{Kangaroo}|\neg \text{Beast}) = 0.0, \]

\[ \text{MaternityKangaroo} \equiv \]
\[ \text{Kangaroo} \cap \exists \text{hasChild.Kangaroo} \]

Inference

\[ P(\text{Parent}(0)|\text{Human}(0)) = 0.232 \]
PDL $\mathcal{CALC}$ can be useful for semantic query extension
PDL can be useful for semantic query extension.

- A probabilistic ontology to model the domain on documents is created.
PDL can be useful for semantic query extension

- A probabilistic ontology to model the domain on documents is created
- Documents are linked to this ontology through indexes
PDL can be useful for semantic query extension

- A probabilistic ontology to model the domain on documents is created
- Documents are linked to this ontology through indexes
- Search procedure
PDL can be useful for semantic query extension

- A probabilistic ontology to model the domain on documents is created
- Documents are linked to this ontology through indexes
- Search procedure
- Query extension
PDL can be useful for semantic query extension

- A probabilistic ontology to model the domain on documents is created
- Documents are linked to this ontology through indexes
- Search procedure
- Query extension
- Ranking results according to their relevance
Search Procedure

Given a CRAŁC ontology:

- Find a set of documents related to the keywords
Search Procedure

Given a CRALC ontology:

- Find a set of documents related to the keywords
- Concepts and roles are found through indexes $\rightarrow$ evidence
Search Procedure

Given a CRALC ontology:

- Find a set of documents related to the keywords
- Concepts and roles are found through indexes → evidence
- A relational Bayesian network propositionalized is built
Given a CRALC ontology:

- Find a set of documents related to the keywords
- Concepts and roles are found through indexes → evidence
- A relational Bayesian network propositionalized is built
Extending the Query

- Ontology provides terms that may be added to the query
Extending the Query

- Ontology provides terms that may be added to the query
- Inference is performed in the relational Bayesian network during search
Extending the Query

- Ontology provides terms that may be added to the query.
- Inference is performed in the relational Bayesian network during search.
- Probability of all concepts that are not evidence in the RBN is inferred.
Extending the Query

- Ontology provides terms that may be added to the query
- Inference is performed in the relational Bayesian network during search
- Probability of all concepts that are not evidence in the RBN is inferred
- Concepts with highest probabilities are input for the ranking results phase
Ranking Procedure

- Documents related to concepts are retrieved and ranked according to their probability
Ranking Procedure

- Documents related to concepts are retrieved and ranked according to their probability.
- These documents are shown together with documents initially selected.
Ranking Procedure

- Documents related to concepts are retrieved and ranked according to their probability
- These documents are shown together with documents initially selected
- Merged ordered list is exhibited
The Lattes Curriculum Platform

J. E. Ochoa Luna (POLI-USP)
The Ontology

Researcher ≡ Person
\( \sqcap (\exists \text{hasPublication}.\text{Publication} \sqcap \exists \text{hasSupervision}.\text{Supervision} \sqcap \exists \text{hasParticipation}.\text{Board}) \)

\( P(\text{NearCollaborator}) \mid \text{Researcher} \sqcap \exists \text{sharePublication.}\exists \text{hasSameInstitution.}\text{Researcher} \)

FacultyNearCollaborator ≡ NearCollaborator
\( \sqcap \exists \text{sameExaminationBoard.}\text{Researcher} \)

\( P(\text{NullMobilityResearcher}) \mid \text{Researcher} \sqcap \exists \text{wasAdvised.} \exists \text{hasSameInstitution.}\text{Researcher} \) = 0.95

StrongRelatedResearcher ≡ Researcher
\( \sqcap (\exists \text{sharePublication.}\text{Researcher} \sqcap \exists \text{wasAdvised.}\text{Researcher}) \)

InheritedResearcher ≡ Researcher
\( \sqcap (\exists \text{sameExaminationBoard.}\text{Researcher} \sqcap \exists \text{wasAdvised.}\text{Researcher}) \)
Goal

Mapping researchers in Bayesian networks
Query Results

Goal

Mapping researchers in Bayesian networks

“Bayesian networks”
Indexing allow us to instantiate properties where the query occurs → propositionalization
Query Extension

- Indexing allow us to instantiate properties where the query occurs → propositionalization
- Researcher(0) contains “Bayesian networks” in a publication → hasPublication(0, 1) is set to true
Indexing allow us to instantiate properties where the query occurs → propositionalization

Researcher(0) contains “Bayesian networks” in a publication → hasPublication(0, 1) is set to true

Related concepts lead to extensions of the original query → sharePublication(0, 2)
Final Result

Semantic Query Extension
Results merged
<table>
<thead>
<tr>
<th>Name</th>
<th>Selected publications</th>
<th>Supervised works</th>
<th>Board participations</th>
<th>Strong related researchers</th>
<th>Near collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. G. C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. B. de M.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. C. F. O.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. C. F. da R.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. T. R.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. F. M.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Preliminary Qualitative Analysis

- Focus on searching researchers that best match several topics
Preliminary Qualitative Analysis

- Focus on searching researchers that best match several topics
- 1964 documents were considered
Preliminary Qualitative Analysis

- Focus on searching researchers that best match several topics
- 1964 documents were considered
- 20 topics evaluated: “Pattern recognition”, “Probabilistic logic”, “Bayesian networks” and so on
Preliminary Qualitative Analysis

- Focus on searching researchers that best match several topics
- 1964 documents were considered
- 20 topics evaluated: “Pattern recognition”, “Probabilistic logic”, “Bayesian networks” and so on
- Semantic information retrieval analysis is still an open issue
Conclusions

- A mix of web documents and probabilistic ontologies
Conclusions

- A mix of web documents and probabilistic ontologies
- Two basic steps: a) a probabilistic ontology is constructed b) search for instance concepts that best match user queries
Conclusions

- A mix of web documents and probabilistic ontologies
- Two basic steps: a) a probabilistic ontology is constructed b) search for instance concepts that best match user queries
- Preliminary results have focused a real-world domain — Lattes scientific repository
Future Work

- Investigate the scalability of our methods
Future Work

- Investigate the scalability of our methods
- Further experiments
The End

Thank you