

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

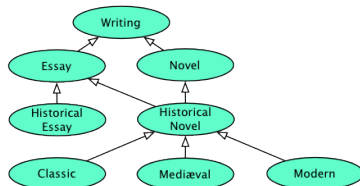


Outline

- 1 Introduction
- 2 Probabilistic Description Logic *CRALL*
- 3 Semantic Query Extension with *CRALL*
- 4 Preliminary Results
- 5 Conclusions
- 6 Future Work



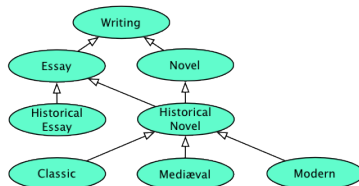
Semantic Information Retrieval



- Focus: use of ontologies to improve keyword-based search



Semantic Information Retrieval



- 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) *CRALLC*



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 $CRALC$

- $CRALC$ is a probabilistic extension of the DL ALC .



Probabilistic Description Logic CRALC

- CRALC is a probabilistic extension of the DL ALC .
- The following constructors are available in ALC : *conjunction* ($C \sqcap D$), *disjunction* $C \sqcup D$, *negation* ($\neg C$), *existential restriction* ($\exists r.C$), and *value restriction* ($\forall r.C$).



Probabilistic Inclusions and their Semantics

- $P(A|B) = \alpha$



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

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



Example

$$P(\text{Animal}) = 0.9,$$

$$P(\text{Rational}) = 0.6,$$

$$P(\text{hasChild}) = 0.3,$$

Human \equiv Animal \sqcap Rational,

Beast \equiv Animal \sqcap \neg Rational,

Parent \equiv

Human \sqcap \exists hasChild.Human,

$$P(\text{Kangaroo}|\text{Beast}) = 0.4,$$

$$P(\text{Kangaroo}|\neg\text{Beast}) = 0.0,$$

MaternityKangaroo \equiv

Kangaroo \sqcap \exists hasChild.Kangaroo



Example

$P(\text{Animal}) = 0.9,$

$P(\text{Rational}) = 0.6,$

$P(\text{hasChild}) = 0.3,$

$\text{Human} \equiv \text{Animal} \sqcap \text{Rational},$

$\text{Beast} \equiv \text{Animal} \sqcap \neg \text{Rational},$

$\text{Parent} \equiv$

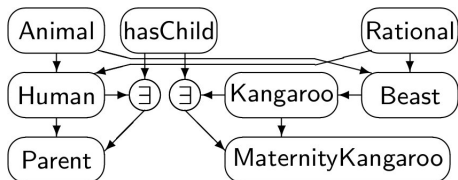
$\text{Human} \sqcap \exists \text{hasChild}.\text{Human},$

$P(\text{Kangaroo}|\text{Beast}) = 0.4,$

$P(\text{Kangaroo}|\neg \text{Beast}) = 0.0,$

$\text{MaternityKangaroo} \equiv$

$\text{Kangaroo} \sqcap \exists \text{hasChild}.\text{Kangaroo}$



Example

$$P(\text{Animal}) = 0.9,$$

$$P(\text{Rational}) = 0.6,$$

$$P(\text{hasChild}) = 0.3,$$

$\text{Human} \equiv \text{Animal} \sqcap \text{Rational},$

$\text{Beast} \equiv \text{Animal} \sqcap \neg \text{Rational},$

$\text{Parent} \equiv$

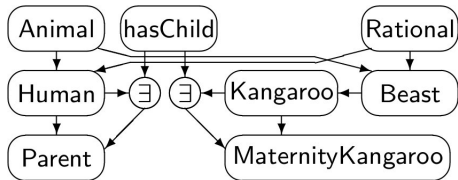
$\text{Human} \sqcap \exists \text{hasChild}.\text{Human},$

$$P(\text{Kangaroo}|\text{Beast}) = 0.4,$$

$$P(\text{Kangaroo}|\neg \text{Beast}) = 0.0,$$

$\text{MaternityKangaroo} \equiv$

$\text{Kangaroo} \sqcap \exists \text{hasChild}.\text{Kangaroo}$



Inference

$$P(\text{Parent}(0)|\text{Human}(0)) = 0.232$$



Proposal

PDL *CRALC* can be useful for semantic query extension



Proposal

PDL *CRALC* can be useful for semantic query extension

- A probabilistic ontology to model the domain on documents is created



Proposal

PDL *CRALC* 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



Proposal

PDL *CRALC* 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



Proposal

PDL *CRALLC* 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



Proposal

PDL *CRALLC* 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 *CRALLC* 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 → evidence



Search Procedure

Given a *CRALC* ontology:

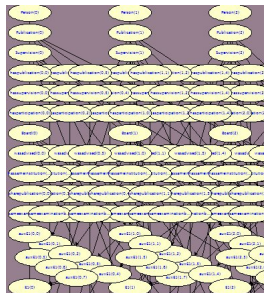
- Find a set of documents related to the keywords
- Concepts and roles are found through indexes → evidence
- A relational Bayesian network propositionalized is built



Search Procedure

Given a *CRALLC* 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

The screenshot shows a web browser window titled "Curriculum System of Curriculum Lattes (José Eduardo Ochoa Luna) - Mozilla Firefox". The address bar contains the URL <http://buscatextual.cnpq.br/buscatextual/visualizacv.jsp?id=K4735082D9&idiomaExibicao=2>. The page content includes a profile for José Eduardo Ochoa Luna, a "Certified by CNPQ" badge, and two expandable sections: "Personal Information" and "Formal Education/Degrees".

Personal Information:

- Name:** José Eduardo Ochoa Luna
- Biographic Citation:** OCHOA, J. E.
- Gender:** Male
- Professional Address:** Universidade de São Paulo, Escola Politécnica, Av. Prof. Mello Moraes, 2231, Cidade Universitária, 05508-220 - São Paulo, SP - Brazil. Phone: (11) 30915127. Website URL: <http://www.pmr.usp.usp.br/>

Formal Education/Degrees:

- 2006:** Ph.D. in progress in (Concurso CAPES 5) - Universidade de São Paulo, USP. Thesis Advisor: Pablo Dagliardi Cosman. Granting of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, ... Keywords: Logica Probabilística; Redes Superiores. Major Area: Exact and Earth Sciences / Area: Computer Science / Subarea: Sistemas de Computação / Specialty: Inteligência Artificial. Address Sector: Information Technology.

The Ontology

Researcher \equiv	Person $\sqcap (\exists \text{hasPublication. Publication}$ $\sqcap \exists \text{hasSupervision. Supervision} \sqcap \exists \text{hasParticipation. Board})$
$P(\text{NearCollaborator})$	$ \text{Researcher} \sqcap \exists \text{sharePublication.} \exists \text{hasSameInstitution.}$ $\text{.Researcher}) = 0.95$
FacultyNearCollaborator \equiv	NearCollaborator $\sqcap \exists \text{sameExaminationBoard. Researcher}$
$P(\text{NullMobilityResearcher})$	$ \text{Researcher} \sqcap \exists \text{wasAdvised.}$ $\exists \text{hasSameInstitution. Researcher}) = 0.98$
StrongRelatedResearcher \equiv	Researcher $\sqcap (\exists \text{sharePublication. Researcher} \sqcap$ $\exists \text{wasAdvised. Researcher})$
InheritedResearcher \equiv	Researcher $\sqcap (\exists \text{sameExaminationBoard. Researcher} \sqcap$ $\exists \text{wasAdvised. Researcher})$



Query Results

Goal

Mapping researchers in Bayesian networks



Query Results

Goal

Mapping researchers in Bayesian networks

“Bayesian networks”

Document	Summary
Currículo do Sistema de Currículos Lattes (I. B. de M.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (F. T. R.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (F. G. C.)	+ Mostrar info Formação cor
Currículo do Sistema de Currículos Lattes (C. P. de C.)	+ Mostrar info Formação cor
Currículo do Sistema de Currículos Lattes (G. A. G.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (J. C. F. da R.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (J. E. O. L.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (M. C. C.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (L. W.)	+ Mostrar info Formação cor
Currículo do Sistema de Currículos Lattes (C. E. T.)	+ Mostrar info Idiomas Atuaç
Currículo do Sistema de Currículos Lattes (A. K.)	+ Mostrar info Formação cor



Query Extension

- 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



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



Final Result

Semantic
Query
Extension
Results merged



Final Result

Semantic
Query
Extension
Results merged

Query results for 'bayesian networks'

[E. G. C.](#)

[Selected publications](#)
[Supervised works](#)
[Board participations](#)
[Strong related researchers](#)
[Near collaborators](#)

[I. B. de M.](#)

[Board participations](#)
[Strong related researchers](#)
[Near collaborators](#)

[A. C. F. O.](#)

[Selected publications](#)
[Strong related researchers](#)
[Near collaborators](#)

[J. C. F. da R.](#)

[Selected publications](#)
[Strong related researchers](#)
[Near collaborators](#)

[E. T. R.](#)

[Strong related researchers](#)
[Near collaborators](#)

[P. E. M.](#)

[Strong related researchers](#)



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: “Patern 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: “Patern 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

