Semantic Query Extension through Probabilistic Description Logics

José Ochoa Luna[1] Kate Revoredo[2] Fabio Cozman[1]

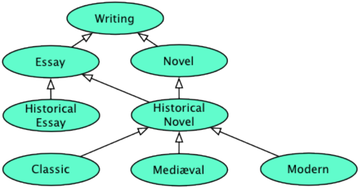
Decision Making Lab.
Escola Politécnica - Universidade de São Paulo

Departamento de Informática Aplicada
Unirio

URSW 2010

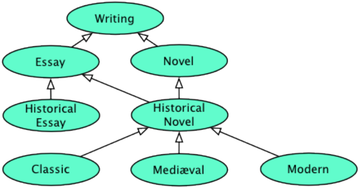
Outline

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 $\rightarrow$ uncertainty $\rightarrow$ (PDL) CR$\mathcal{ALC}$

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(Concept|Query)

Probabilistic Description Logic CR$\mathcal{ALC}$

- CR$\mathcal{ALC}$ is a probabilistic extension of the DL $\mathcal{ALC}$.

Probabilistic Description Logic CR$\mathcal{ALC}$

- CR$\mathcal{ALC}$ is a probabilistic extension of the DL $\mathcal{ALC}$.
- The following constructors are available in $\mathcal{ALC}$: *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

- $P(\mathsf{A}|\mathsf{B}) = \alpha$

Probabilistic Inclusions and their Semantics

- $P(\mathsf{A}|\mathsf{B}) = \alpha$
- $\forall x \in \mathcal{D} \;:\; P(\mathsf{A}(\mathsf{x})|\mathsf{B}(\mathsf{x})) = \alpha$

Probabilistic Inclusions and their Semantics

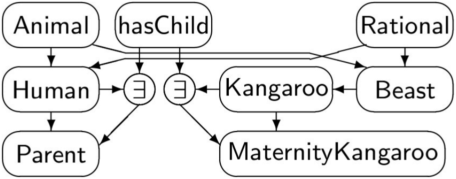
- $P(\mathrm{A}|\mathrm{B}) = \alpha$
- $\forall x \in \mathcal{D} \;:\; P(\mathrm{A(x)}|\mathrm{B(x)}) = \alpha$
- $P(\mathrm{Professor(Maria)}|\mathrm{Researcher(Maria)}) = 0.4$

Example

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

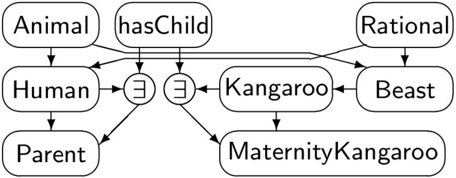
Example

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

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



Inference

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

Proposal

PDL CR$\mathcal{ALC}$ can be useful for semantic query extension

Proposal

PDL $\text{CR}\mathcal{ALC}$ can be useful for semantic query extension

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

Proposal

PDL CR$\mathcal{ALC}$ 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 $\textsc{cr}\mathcal{ALC}$ 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 CR$\mathcal{ALC}$ 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 CR$\mathcal{ALC}$ 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 CR$\mathcal{ALC}$ ontology:

- Find a set of documents related to the keywords

Search Procedure

Given a CR$\mathcal{ALC}$ ontology:

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

Search Procedure

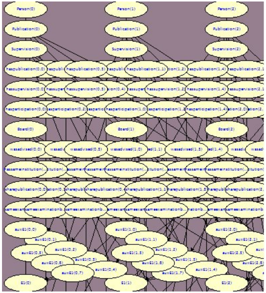
Given a CR$\mathcal{ALC}$ ontology:

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

Search Procedure

Given a CR$\mathcal{ALC}$ ontology:

- Find a set of documents related to the keywords
- Concepts and roles are found through indexes $\rightarrow$ evidence
- A relational Bayesian network propositionalizated 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

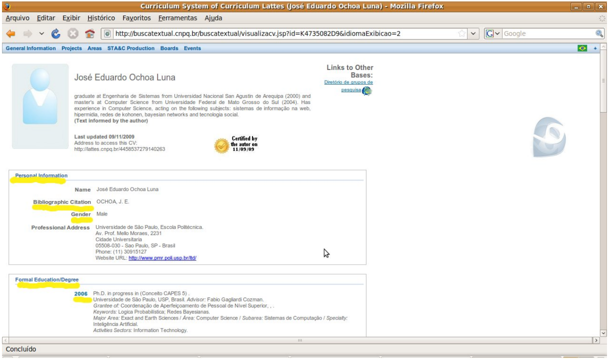
Ranking Procedure

- 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 Ontology

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

Query Results

Goal

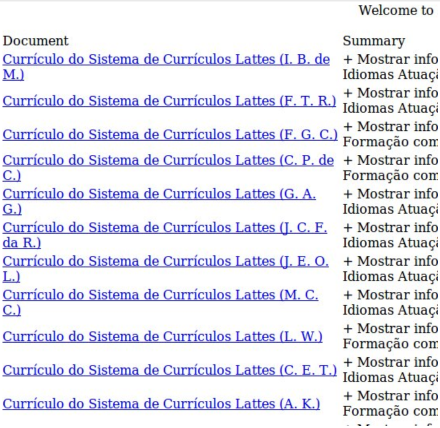
Mapping researchers in Bayesian networks

Query Results

Goal

Mapping researchers in Bayesian networks

"Bayesian networks"



Welcome to

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

Query Extension

- Indexing allow us to instantiate properties where the query occurs $\longrightarrow$ propositionalization

Query Extension

- Indexing allow us to instantiate properties where the query occurs $\rightarrow$ propositionalization
- Researcher(0) contains "Bayesian networks" in a publication $\rightarrow$ $\text{hasPublication}(0, 1)$ is set to true

Query Extension

- Indexing allow us to instantiate properties where the query occurs $\rightarrow$ propositionalization
- Researcher(0) contains "Bayesian networks" in a publication $\rightarrow$ $\text{hasPublication}(0, 1)$ is set to true
- Related concepts lead to extensions of the original query $\rightarrow$ $\text{sharePublication}(0, 2)$

Final Result

Semantic Query Extension

Results merged

Final Result

Semantic Query Extension

Results merged

Query results for 'bayesian networks'

F. 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

F. 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