A STUDY AT THE CROSSROAD OF ARCHIVAL DESCRIPTION, LOGIC AND ONTOLOGIES
GOAL: LEVERAGING LOGIC-BASED AI FOR USER QUERIES

Vague query in mind...

User Interface / Visualization

We are here

AI assistant / "Cognitive" facilitator

RiC-O Ontology
BASIC NOTIONS OF LOGIC-BASED REASONING
BASIC NOTIONS OF LOGIC-BASED REASONING: RICO EXAMPLE

SPARQL Query
To try out some SPARQL queries against the selected dataset, enter your query here.

Example Queries

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX refs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rico: <https://www.ica.org/standards/RIIC/ontology#>

select ?s ?o WHERE {
  ?s rico:isAssociatedWithPlace ?o
}
LIMIT 50000
```

Content Type (SELECT)
- JSON

Content Type (GRAPH)
- Turtle
BASIC NOTIONS OF LOGIC-BASED REASONING: RICO EXAMPLE
Leveraging Machine learning, Logic and linguistics
LEVERAGING ML, LOGIC AND LINGUISTICS: ILP

- Inductive Logic Programming (Logic Machine Learning)
LEVERAGING ML, LOGIC AND LINGUISTICS

▸ Inductive Logic Programming (Logic Machine Learning)
  ▸ Good rules are persisted into the theory
  ▸ And can be used to deduce new knowledge

```prolog
animal(A).
fly(A) :-
  class(A,mammal).
fly(A) :-
  has_milk(A).
fly(A) :-
  homeothermic(A).

[clauses constructed] [6]
[search time] [0.0001399999999793]

[best clause]
fly(bat).
[pos cover = 1 neg cover = 0] [posonly bayes estimate] [0.0918424535831607]
[atoms left] [0]
[positive examples left] [0]
[estimated time to finish (secs)] [0.01]

[theory]

[Rule 1] [Pos cover = 4 Rand cover = 6]
fly(A) :-
  class(A,bird).

[Rule 2] [Pos cover = 1 Rand cover = 1]
fly(bat).

[Training set performance]

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred</td>
<td>+5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Accuracy = 1

[Training set summary] [[5,0,0,0]]
[time taken] [0.002160999999999963]
[total clauses constructed] [20]
Updating rules file...
true.

?- dynamic(class/2).
true.

?- class
LEVERAGING ML, LOGIC AND LINGUISTICS: ILP APPLIED TO RICO

- Inductive Logic Programming (Logic Machine Learning)
  - Good rules are persisted into the theory
  - And can be used to deduce new knowledge

Theory (Rules) + Knowledge Base → Induction → Inferring facts

- $\text{skos\_broader}(A, B) :- \text{skos\_narrower}(B, A)$.
- $\text{skos\_narrower}(A, B) :- \text{skos\_broader}(B, A)$.
- $\text{rico\_activityIsSourceOfPerformanceRelation}(A, B) :- \text{rico\_performanceRelationHasSource}(B, A)$.
- $\text{rico\_agentOriginationRelationHasTarget}(A, B) :- \text{rico\_agentIsTargetOfAgentOriginationRelation}(B, A)$.
- $\text{rico\_describesOrDescribed}(A, B) :- \text{rico\_isOrWasDescribedBy}(B, A)$.
- $\text{rico\_hasDerivedInstantiation}(A, B) :- \text{rico\_isDerivedFromInstantiation}(B, A)$.
- $\text{rico\_hasInstantiation}(A, B) :- \text{rico\_isInstantiationOf}(B, A)$.
- $\text{rico\_isDerivedFromInstantiation}(A, B) :- \text{rico\_hasDerivedInstantiation}(B, A)$.
- $\text{rico\_isInstantiationOf}(A, B) :- \text{rico\_hasInstantiation}(B, A)$.
- $\text{rico\_isOrWasAgentNameOf}(A, B) :- \text{rico\_hasOrHadAgentName}(B, A)$.
- $\text{rico\_isOrWasConstituentOf}(A, B) :- \text{rico\_hasOrHadConstituent}(B, A)$.
- $\text{rico\_isOrWasDescribedBy}(A, B) :- \text{rico\_describesOrDescribed}(B, A)$. 
LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT

- Inductive Logic Programming (Logic Machine Learning)
- Research domain enrichment of the theory
LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT

- Inductive Logic Programming (Logic Machine Learning)
- Research domain enrichment of the theory
  - Ex: a user is asking for records in the diocese of Ascoli.
LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT

- Research domain enrichment of the theory
  - Ex: a user is asking for records in the diocese of Ascoli.

Here is a list of concepts. Which is the closest to the concept of diocese?

```
['base_placeType/amC3%A9nagement%20ou%20construction
', 'base_placeType/ancienne%20commune
', 'base_placeType/arrondissement%20municipal
', 'base_placeType/commune
', 'base_placeType/paroisse
', 'base_placeType/voie%20urbaine
'].
```

Among the concepts you provided, the closest concept to the diocese is `base_placeType/paroisse`. A diocese is a territorial division within the Christian Church, and a parish (paroisse) is a subdivision within a diocese. Parishes are typically smaller administrative units that are overseen by a priest or pastor and serve as local congregations.
LEVERAGING ML, LOGIC AND LINGUISTICS: METATHEORY

- Inductive Logic Programming (Logic Machine Learning)
- Research domain enrichment of the theory
- Explicit definition of the metatheory
LEVERAGING ML, LOGIC AND LINGUISTICS: METATHEORY

SPARQL Query
To try out some SPARQL queries against the selected dataset, enter your query here.

Example Queries

```sparql
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rico: <https://www.ica.org/standards/RIC/ontology#>

SELECT ?s WHERE
 {?s rico:isAssociatedWithPlace ?o}
LIMIT 50000
```

Table: Response 16821 results in 2.111 seconds

<table>
<thead>
<tr>
<th>s</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://data.archives-nationales.culture.gouv.fr/fr/place/15%2C%20rue%20des%20Saints-Pl%C3%A8ves%2C%2075006">http://data.archives-nationales.culture.gouv.fr/fr/place/15%2C%20rue%20des%20Saints-Pl%C3%A8ves%2C%2075006</a>...</td>
<td><a href="http://data.archives-nationales.culture.gouv.fr/fr/place/15%2C%20rue%20des%20Saints-Pl%C3%A8ves%2C%2075006">http://data.archives-nationales.culture.gouv.fr/fr/place/15%2C%20rue%20des%20Saints-Pl%C3%A8ves%2C%2075006</a>...</td>
</tr>
</tbody>
</table>
The concept of a *lost record* becomes an “Impossible concept”
homeothermic(A).
[clauses constructed] [5]
[search time] [0.0001809999999998673]
[best clause]
fly(bat).
[pos cover = 1 neg cover = 0] [posonly bayes estimate] [0.09104245335831687]
[atoms left] [0]
[positive examples left] [0]
[estimated time to finish (secs)] [0.0]

[theory]

[Rule 1] [Pos cover = 4 Rand cover = 5]
fly(A) :-
  class(A,bird).

[Rule 2] [Pos cover = 1 Rand cover = 1]
fly(bat).

[Training set performance]

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ 5</td>
</tr>
<tr>
<td></td>
<td>- 0</td>
</tr>
<tr>
<td>Pred</td>
<td>0 0 0</td>
</tr>
<tr>
<td></td>
<td>5 0 5</td>
</tr>
</tbody>
</table>

Accuracy = 1
[Training set summary] [[5,0,0,0]]
[t ime taken] [0.0627130008000000083]
[total clauses constructed] [20]
Updating rules file...

?- dynamic(class/2).
true.

?- I X
LEVERAGING ML, LOGIC AND LINGUISTICS: INTENSIONAL LOGIC

- Inductive Logic Programming (Logic Machine Learning)
- Research domain enrichment of the theory
- Explicit definition of the metatheory
- Intensionalization of the metatheory
LEVERAGING ML, LOGIC AND LINGUISTICS: INTENSIONAL LOGIC

- Inductive Logic Programming (Logic Machine Learning)
- Research domain enrichment of the theory
- Explicit definition of the metatheory
- Intensionalization of the metatheory

Logic-based AI capability for archival ontologies
A typical Use Case

"I am looking for a correspondence on theology between Descartes and Galileo."

"No, I am specifically interested in Descartes and Galileo."

"I cannot find any such correspondence. Are you flexible on one of the agents involved in the correspondence? For instance, I found some correspondence on Theology between Galileo and Mersenne."
A typical Use Case

"No, I am specifically interested in Descartes and Galileo."

"Sure. Would you consider a mention of Descartes or Galileo in a letter as relevant?"

"Absolutely."
A typical Use Case

"Absolutely."

"No, I do care that both agents are authors."

"Would you like both agents to be authors or is acceptable if one of them is simply mentioned? For example, I found numerous mentions of Galileo in the correspondance between Descartes and Elizabeth."

RiCO-AI
A typical Use Case
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
Intensional Logic: A network of concepts
A COLLABORATIVE PROJECT

› Partnership:
  › InterPARES-Trust-AI (Vancouver, BC)
  › SCAI - Sorbonne Center for Artificial Intelligence (Paris, France)
  › ICA - International Council on Archives / EGAD - Expert Group on Archival Description
  › ANF - Archives Nationales de France (Paris, France)
  › Ecole des Chartes / Centre Jean Mabillon (Paris, France)
  › ObTIC - Observatoire des Textes, Idées et Corpus (Paris, France)

› Collaborators:
DEMONSTRATION: PROTOTYPICAL AI CAPABILITIES IN RICO

▸ Already “some AI” integrated via the RDFS reasoner
▸ Integrate logic-based reasoning into SPARQL queries
  ▸ Resolution
  ▸ Unification
  ▸ Backtracking
▸ Induce theory from ground triples (ground facts)
▸ “Ontology completion” with ChatGPT
METHODOLOGICAL APPROACH

- Agile-like iterative process including theoretical research
  - Progress made on implementation level informs research
  - Research guides implementation
  - Gives us versions of demonstrable prototypes early
  - Avoid scalability issues and technical debt

- No definite specifications but bringing important pieces of a puzzle together
  - We know we need a certain number of important elements: Java interface, Logic programming, Intensional reasoner, RiC-O dataset, SPARQL server,...
  - Agile approach will help us design the product as we’re moving forward
First step: bringing the pieces together
Prototype architecture: bringing the pieces together
Prototype architecture: bringing the pieces together

- Some of those pieces are very small
- It is not a final design / specification
- Experimental (agile approach): some will go away or be rebuilt
ACCOMPLISHMENTS

• With Regis funding, we have hired a Computer Science student as a summer research assistant
• We have succeeded in loading the Archives Nationales de France RiC-O dataset (the largest existing dataset in the RiC-O format) onto an AWS server that we can query remotely.
• We have a skeleton of Java program that can interface with the different services of the project (AWS server via SPARQL, Prolog, Haskell and OpenAI’s gpt3.5-turbo)
• We have a prototype that can learn rules from ground facts in Prolog using ILP (Inductive Logic Programming)
• We have a prototype for implementing an intensional interpretation of those rules (in Haskell)
• We are able to query the SPARQL knowledge base via Prolog
• We have applied to a National Endowment for Humanities (NEH) grant: https://www.neh.gov/grants/odh/digital-humanities-advancement-grants
• Updated project description: https://docs.google.com/document/d/1iIPfhAGTnPMS8a7Z_9QxT69NA7SjGRv7B3bP9ygWlQ/edit?usp=share_link
# Our Team

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Institution/Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hugolin Bergier</td>
<td>Associate Professor in Computer Science</td>
<td>Regis University, SCAI</td>
</tr>
<tr>
<td>Kenneth Thibodeau</td>
<td>Retired director of the ERAP</td>
<td>NARA (retired)</td>
</tr>
<tr>
<td>Florence Clavaud</td>
<td>Archivist, Lab director</td>
<td>ICA/EGAD, ANF, Centre Jean Mabillon</td>
</tr>
<tr>
<td>Pierluigi Feliciati</td>
<td>Archival studies</td>
<td>University of Macerata</td>
</tr>
<tr>
<td>Joe Tennis</td>
<td>Archival studies</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Jean-Pierre Desclés</td>
<td>Retired professor in Logic and Linguistics</td>
<td>Sorbonne University</td>
</tr>
<tr>
<td>Arien Gonzales</td>
<td>Academic Librarian</td>
<td>El Colegio de México</td>
</tr>
<tr>
<td>Édouard Vasseur</td>
<td>Archivist</td>
<td>École Nationale des Chartes, Centre Jean Mabillon</td>
</tr>
<tr>
<td>Motasem Alrahabi</td>
<td>Research Scientist in computational linguistics</td>
<td>SCAI, Observatoire des Textes et Corpus</td>
</tr>
<tr>
<td>Cameron Christner</td>
<td>Undergraduate student in Computer Science</td>
<td>Regis University</td>
</tr>
</tbody>
</table>