INTENSIONAL LOGIC FOR RECORDS-IN-CONTEXTS ONTOLOGY

InterPARES Symposium San Benedetto del Tronto, Italy – July 2023

InterPARES Trust



A STUDY AT THE CROSSROAD OF ARCHIVAL DESCRIPTION, LOGIC AND ONTOLOGIES



GOAL: LEVERAGING LOGIC-BASED AI FOR USER QUERIES



BASIC NOTIONS OF LOGIC-BASED REASONING



BASIC NOTIONS OF LOGIC-BASED REASONING: RICO EXAMPLE



BASIC NOTIONS OF LOGIC-BASED REASONING: RICO EXAMPLE

ry here. Prefixe rat	S		
Prefixe rat	S		
rdf			
	owi xsd		
Content Type (SELECT)		Content Type (GRAPH)	
JSON	~	Turtle	
	Simpl	e view⊂ Filipse⊽ Filter query results	Page size: 50
	JSON	JSON ~	JSON ~ Turtle

BASIC NOTIONS OF LOGIC-BASED REASONING: RICO EXAMPLE

Apache Jena Fuseki 🛢 datasets 🏘 manage 🚯 help				server sta
ricoDatasetInferred				
e query				
SPARQL Query				
o try out some SPARQL queries against the selected dataset, enter your qu	ery here.			
xample Queries		Prefixes		
Selection of triples Selection of classes		rdf rdfs owl xsd		
PARQL Endpoint	Content Type (SELECT)		Content Type (GRAPH)	
/ricoDatasetInferred/query	JSON	~	Turtle	
<pre>PREFIX rdfs: <http: 01="" 2000="" rdf-schema#="" www.w3.org=""> PREFIX skos: <http: 02="" 2004="" core#="" skos="" www.w3.org=""> PREFIX rico: <https: ontology#="" ric="" standards="" www.ica.org=""> select ?s ?o WHERE {? {?s rico:isAssociatedWithPlace ?o} LIMIT 50000</https:></http:></http:></pre>				<
		Simple	e view□ Ellipse⊠ Filter query results	Page size: 50 🟮 🛓
S		≜ o		
1 <http: 010128="" agent="" data.archives-nationales.culture.gouv.fr=""></http:>		<http: data.archives-nationales.cultu<="" td=""><td>re.gouv.fr/place/15%2C%20rue%20des%20Sa</td><td>aints-P%C3%A8res%2C%20750</td></http:>	re.gouv.fr/place/15%2C%20rue%20des%20Sa	aints-P%C3%A8res%2C%20750
2 <http: 010129="" agent="" data.archives-nationales.culture.gouv.fr=""></http:>		<http: data.archives-nationales.cultu<="" td=""><td>re.gouv.fr/place/30%20place%20Denfert-Roc</td><td>hereau%2C%20Paris%20XIVe></td></http:>	re.gouv.fr/place/30%20place%20Denfert-Roc	hereau%2C%20Paris%20XIVe>
3 <http: 010130="" agent="" data.archives-nationales.culture.gouv.fr=""></http:>		<a>http://data.archives-nationales.cultu	re.gouv.fr/place/19%2C%20boulevard%20de%	%20Courcelles>

Leveraging Machine learning, Logic and linguistics



LEVERAGING ML, LOGIC AND LINGUISTICS: ILP

 Inductive Logic Programming (Logic Machine Learning)



		 ~ 0		hugolinbergier — -bash — 130×45
		<u> </u>	Mac-mini:~ hugolinbergier\$ []	
Name	Date Modified \sim	Size	Ki	
test_rdf.pl	Jun 30, 2023 at 11:08 PM	125 KB	Pe	
rdf2pl.pl	Jun 30, 2023 at 8:36 PM	14 KB	Pe	
SymmetricPred.pl	Jun 30, 2023 at 5:48 PM	661 bytes	Pe	
SymmetricPRed	Jun 30, 2023 at 5:38 PM	Zero bytes	De	
goodfile.txt	Jun 29, 2023 at 10:56 PM	1 KB	Pl	
aleph_analysis.pl	Jun 29, 2023 at 6:38 PM	714 bytes	Pe	
bgRules.pl	Jun 29, 2023 at 6:33 PM	1 KB	Pe	
testPersist.pl	Jun 28, 2023 at 7:51 PM	267 bytes	Pe	
📄 LODv2.hs	Jun 21, 2023 at 3:37 AM	8 KB	De	
LODv2.o	May 18, 2023 at 5:20 AM	160 KB	ot	
LODv2.hi	May 18, 2023 at 5:20 AM	7 KB	D¢	
KB2.0	May 18, 2023 at 5:20 AM	93 KB	ot	
📄 KB2.hi 😽	May 18, 2023 at 5:20 AM	4 KB	De	
🛅 random_facts.pl	May 18, 2023 at 12:53 AM	1 KB	Pe	
📄 animals.pl	May 15, 2023 at 10:19 PM	2 KB	Pe	
🍺 queries.hs	May 15, 2023 at 3:44 AM	2 KB	De	
📄 KB2.hs	May 15, 2023 at 3:44 AM	3 KB	De	
📄 animals_aleph_fly.pl	Jun 27, 2023 at 9:59 PM	3 KB	Pe	
💾 auto_ilp.pl	Jun 22, 2023 at 10:39 PM	768 bytes	Pe	
📄 rules.txt	Jun 21, 2023 at 5:14 PM	488 bytes	Pl	
newRules.txt	Jun 21, 2023 at 5:14 PM	38 bytes	PI:	
📄 animals_aleph_lives_in_water.pl	Jun 21, 2023 at 5:14 PM	3 KB	Pe	
📄 animals_aleph_has_gills.pl	Jun 21, 2023 at 5:14 PM	3 KB	Pe	
animals_aleph_homeothermic.pl	Jun 21, 2023 at 5:14 PM	3 KB	Pe	
📄 animals_aleph_lay_eggs.pl	Jun 21, 2023 at 5:14 PM	3 KB	Ρε	
animals_aleph_has_milk.pl	Jun 21, 2023 at 5:14 PM	3 KB	Pe	
🎽 Macintosh HD > 🔯 Users > 🛅 hugolinbergier > 🛅 dev > 🚞	RICO > RICO-AI > auto_ilp.pl	A 1/A		



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



animal(A). class(A,mammal). fly(A) :has_milk(A). homeothermic(A). [clauses constructed] [5] [search time] [0.00017399999999997973] [best clause] fly(bat). [pos cover = 1 neg cover = 0] [posonly bayes estimate] [0.09104245335831607] [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] Actual 5 0 Pred - 0 0 0 5 5 0 Accuracy = 1[Training set summary] [[5,0,0,0]] [time taken] [0.0021609999999999963] [total clauses constructed] [20] Updating rules file... true. ?- dynamic(class/2). true.

?- clas

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



skos_broader(A,B) : skos_narrower(B,A).

skos_narrower(A,B) : skos_broader(B,A).

rico_activityIsSourceOfPerformanceRelation(A,B) : rico_performanceRelationHasSource(B,A).

rico_agentOriginationRelationHasTarget(A,B) : rico_agentIsTargetOfAgentOriginationRelation(B,A).

rico_describesOrDescribed(A,B) : rico_isOrWasDescribedBy(B,A).

rico_hasDerivedInstantiation(A,B) : rico_isDerivedFromInstantiation(B,A).

rico_hasInstantiation(A,B) : rico_isInstantiationOf(B,A).

rico_isDerivedFromInstantiation(A,B) : rico_hasDerivedInstantiation(B,A).

rico_isInstantiationOf(A,B) : rico_hasInstantiation(B,A).

rico_isOrWasAgentNameOf(A,B) : rico_hasOrHadAgentName(B,A).

rico_isOrWasConstituentOf(A,B) : rico_hasOrHadConstituent(B,A).

rico_isOrWasDescribedBy(A,B) : rico_describesOrDescribed(B,A).

LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT



1877 **1**115

LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT



[?- placeTypes(PTL).

PTL = ['\'base_placeType/am%C3%A9nagement%20ou%20construction\'', '\'base_placeType/ancienne%20commun e\'', '\'base_placeType/arrondissement%20municipal\'', '\'base_placeType/commune\'', '\'base_placeTyp e/paroisse\'', '\'base_placeType/quartier\'', '\'base_placeType/voie%20urbaine\''].

1877 IHS

LEVERAGING ML, LOGIC AND LINGUISTICS: DOMAIN ENRICHMENT

?- placeTypes(PTL).

PTL = ['\'base_placeType/am%C3%A9nagement%20ou%20construction\'', '\'base_placeType/ancienne%20commun e\'', '\'base_placeType/arrondissement%20municipal\'', '\'base_placeType/commune\'', '\'base_placeTyp e/paroisse\'', '\'base_placeType/quartier\'', '\'base_placeType/voie%20urbaine\''].

Research and here is a list of concepts. Which is the closest to the concept of diocese? HB domain Natural Language enrichment of [\\base placeType/am%C3%A9nagement%20ou%20construction\". the theory "\'base placeType/ancienne%20commune\", • Ex: a user is '\'base_placeType/arrondissement%20municipal\", '\'base_placeType/commune\", asking for \'base_placeType/paroisse\", '\'base_placeType/quartier\", records in the '\'base placeType/voie%20urbaine\"]. diocese of Ascoli. 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



LEVERAGING ML, LOGIC AND LINGUISTICS: METATHEORY



LEVERAGING ML, LOGIC AND LINGUISTICS: METATHEORY

1877 IHS

Apache Jena Fuseki 🗧 datasets 🍫 manage 🚯 help							serv	ver statu:
icoDatasetInferred								
Q query								
SPARQL Query								
o try out some SPARQL queries against the selected dataset, enter your q	uery here.							
Example Queries		Prefixes						
Selection of triples Selection of classes		rdf rdfs owl xsd						
PARQL Endpoint	Content Type (SELECT)			Content Type	(GRAPH)			
/ricoDatasetInferred/query	JSON	~	-	Turtle				~
<pre>1 v PREFIX rdf: <http: 02="" 1999="" 22-rdf-syntax-ns#="" www.w3.org=""> 2 PREFIX rdfs: <http: 01="" 2000="" rdf-schema#="" www.w3.org=""> 3 PREFIX skos: <http: 02="" 2004="" core#="" skos="" www.w3.org=""> 4 PREFIX rico: <https: ontology#="" ric="" standards="" www.ica.org=""> 5 6 select ?s ?o WHERE 7 {?s rico:isAssociatedWithPlace ?o} 8 LIMIT 50000</https:></http:></http:></http:></pre>							<	
Table E Response 18821 results in 2.111 seconds s		Sin	nple	view⊡ Ellipset	2 Filter query results	Page size:	50 📀	76
<pre>1 <http: 010128="" agent="" data.archives-nationales.culture.gouv.fr=""></http:></pre>		<http: data.archives-nationales.c<="" td=""><td>culture</td><td>e.gouv.fr/place/1</td><td>5%2C%20rue%20des%20Sa</td><td>aints-P%C3%A8r</td><td>es%2C%2</td><td>075006.</td></http:>	culture	e.gouv.fr/place/1	5%2C%20rue%20des%20Sa	aints-P%C3%A8r	es%2C%2	075006.
2 <http: 010129="" agent="" data.archives-nationales.culture.gouv.fr=""></http:>		<http: data.archives-nationales.c<="" td=""><td>ulture</td><td>e.gouv.fr/place/3</td><td>0%20place%20Denfert-Roc</td><td>hereau%2C%20P</td><td>aris%20X</td><td>IVe></td></http:>	ulture	e.gouv.fr/place/3	0%20place%20Denfert-Roc	hereau%2C%20P	aris%20X	IVe>
3 <http: 010130="" agent="" data.archives-nationales.culture.gouv.fr=""></http:>		<http: data.archives-nationales.c<="" td=""><td>ulture</td><td>e.gouv.fr/place/1</td><td>9%2C%20boulevard%20de%</td><td>620Courcelles></td><td></td><td></td></http:>	ulture	e.gouv.fr/place/1	9%2C%20boulevard%20de%	620Courcelles>		



LEVERAGING ML, LOGIC AND LINGUISTICS: METATHEORY

noInstantiation(X) : ??(fusekiInfer,
 (rdf(X, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', 'https://www.ica.org/standards/RiC/ontology#Record'),
 \+ rdf(X, 'https://www.ica.org/standards/RiC/ontology#hasInstantiation', Y))
).



%Rule

rdf(X, 'https://www.ica.org/standards/RiC/ontology#hasInstantiation', _)) : rdf(X, 'http://www.w3.org/1999/02/22-rdf-syntax-ns#type', 'https://www.ica.org/standards/RiC/ontology#Record').

The concept of a *lost record* becomes an "Impossible concept"

	💿 😑 🛑 resources — swipl — 101×39
Name	homeothermic(A). [clauses constructed] [5] [search time] [0.000180999999998673] [best clause]
v in resources	fly(bat).
🖉 📄 rules.txt	[pos cover = 1 neg cover = 0] [posonly bayes estimate] [0.0910424535851007] [atoms left] [0]
newRules.txt	[positive examples left] [0]
newinstanceRule.pl	[estimated time to Finish (secs)] [0.0]
📄 rdfFuseki.pl	[theory]
test_rdf.pl	[Rule 1] [Pos cover = 4 Rand cover = 5]
rdf2pl.pl	fly(A) :-
SymmetricPred.pl	class(A, bird).
SymmetricPRed	[Rule 2] [Pos cover = 1 Rand cover = 1]
goodfile.txt	fly(bat).
aleph_analysis.pl	[Training set performance]
bgRules.pl	Actual
testPersist.pl	+ 5 0 5
LODv2.hs	Pred
LODv2.o	-0000
LODv2.hi	5 0 5
KB2.o	Accuracy = 1
KB2.hi	[Training set summary] [[5,0,0,0]]
random_facts.pl	[time taken] [0.002173000000000083]
animals.pl	Updating rules file
🧃 queries.hs	true.
j≩ KB2.hs	[?- dvnamic(class/2).
animals_aleph_fly.pl	true.
auto_ilp.pl	2-
rules.txt	
newRules.txt	
Manimale along lives in water of	
Macinto 7 10 7 10 7 10 7 10 7 10 7 10 7 10 7 1	

LEVERAGING ML, LOGIC AND LINGUISTICS: INTENSIONAL LOGIC



LEVERAGING ML, LOGIC AND LINGUISTICS: INTENSIONAL LOGIC





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











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







"There is no instance of a letter





























Interactions in a group of people

utual interaction in a group of people Intellectual topic

extension







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:



InterPARES

Trust -







EGAD (BICA



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

•







- 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 OpenAl'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: <u>https://www.neh.gov/grants/odh/digital-humanities-advancement-grants</u>
- Updated project description: <u>https://docs.google.com/document/d/1ilPfhAGTnNPMS8a7Z_9QxT69NA7SjGRv7B3bP9ygWIQ/edit?us</u> <u>p=share_link</u>



OUR TEAM

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