DistEL: A Distributed $\mathcal{EL}^+$ Ontology Classifier

Raghava Mutharaju

Pascal Hitzler

Prabhaker Mateti

Department of Computer Science

Wright State University, Dayton, OH, USA

Setting

$\mathcal{EL}^+$ is a fragment of OWL EL

e.g., SNOMED

and tractable (polytime)

We investigate distributed memory reasoning with $\mathcal{EL}^+$.

Distributed memory reasoning is generally very hard.

Evaluation data

Ontology	#Logical Axioms	#Concepts	#Roles
Not-Galen	8,015	4,242	413
GO	28,897	20,465	1
NCI	46,870	27,653	70
SNOMED	1,038,481	433,106	62
SNOMED-DUP-2	2,076,962	866,212	124
SNOMED-DUP-3	3,115,443	1,299,318	186
SNOMED-GALEN-GO	1,075,393	456,319	476

Table 2. Sizes of (normalized) ontologies we used

Results: single memory reasoners

Ontology	Pellet	jCEL	ELK
Not-Galen	12.0	3.0	1.0
GO	5.0	5.0	2.0
NCI	6.0	7.0	3.0
SNOMED	1,845.0	327.0	24.0
SNOMED-DUP-2	OutOfMemory	687.0	64.0
SNOMED-DUP-3	OutOfMemory	1149.0	93.0
SNOMED-GALEN-GO	OutOfMemory	TIME OUT	TIME OUT

Table 3. Classification time of ontologies using Pellet, jCEL and ELK

Results: DistEL load times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	6.76	6.44	6.67	6.67	7.09
GO	11.58	11.65	11.74	12.59	12.51
NCI	21.13	21.57	21.53	22.15	22.80
SNOMED	382.77	385.09	392.09	398.07	393.57
SNOMED-DUP-2	774.34	767.85	787.10	798.58	826.50
SNOMED-DUP-3	2,160.00	2,160.00	2,113.57	2,194.80	2,233.12
SNOMED-GALEN-GO	—	—	—	—	411.72

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	6.78	6.83	6.74	6.77
GO	12.30	12.46	12.87	12.93
NCI	22.53	22.63	22.66	22.15
SNOMED	396.66	405.94	410.07	412.39
SNOMED-DUP-2	803.43	805.81	828.55	828.78
SNOMED-DUP-3	2,177.13	2315.19	2163.17	2257.94
SNOMED-GALEN-GO	416.99	418.99	419.58	428.43

Table 4. Load times (in seconds) of DistEL

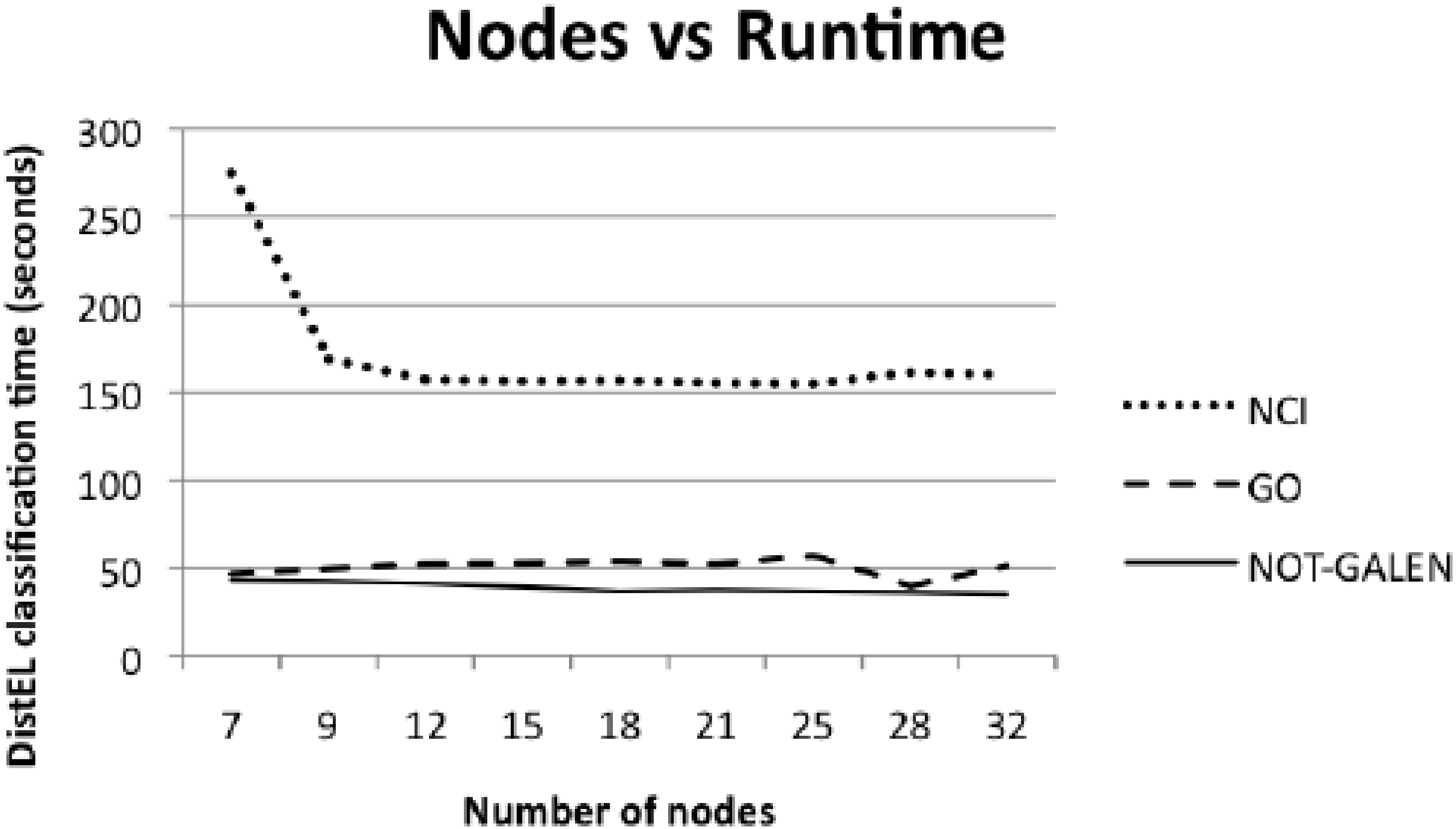
Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

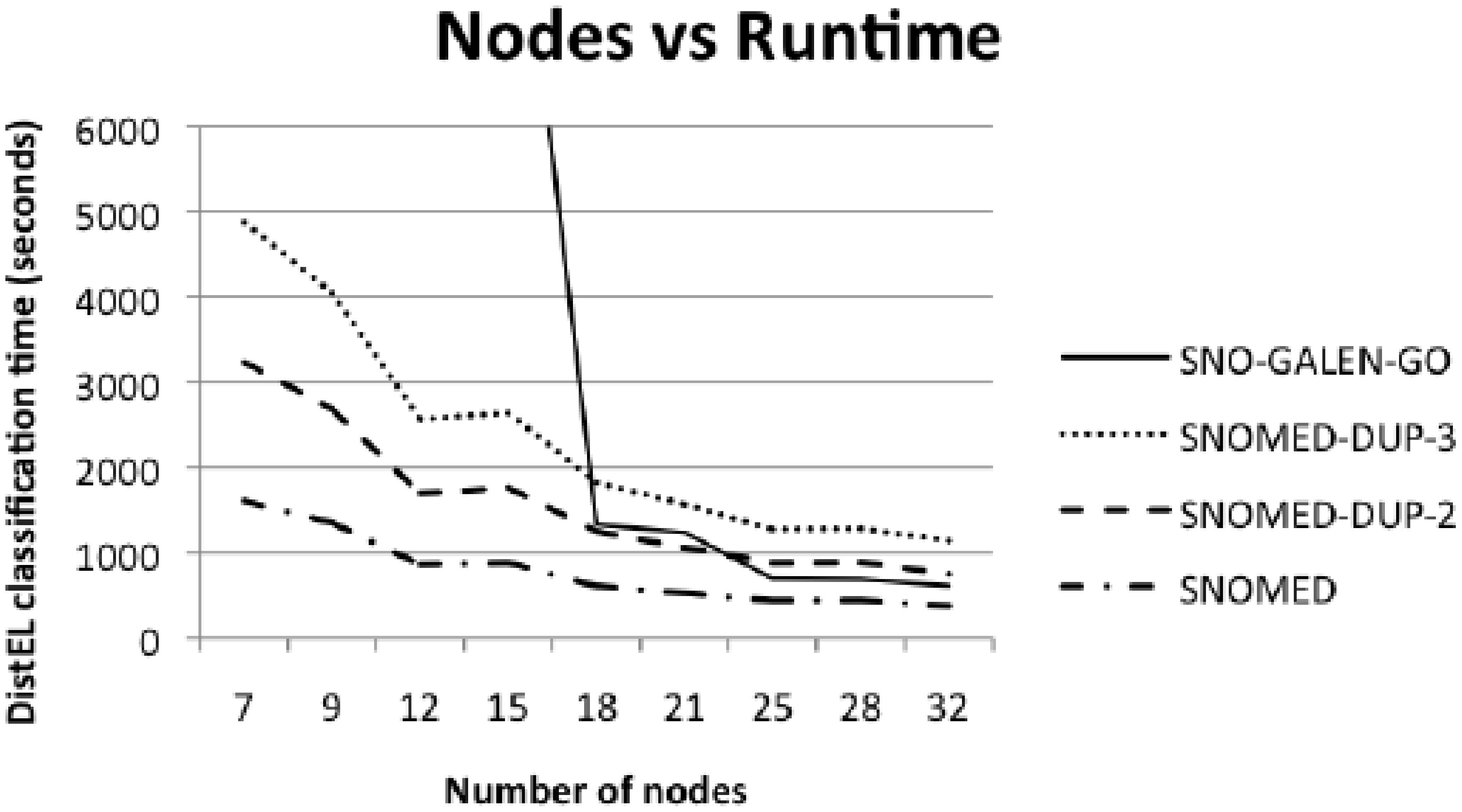
Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

DistEL reasoning: small ontologies



DistEL reasoning: large ontologies



Results: DistEL reasoning times

WRIGHT STATE UNIVERSITY

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

Results: DistEL reasoning times

Ontology	7 nodes	9 nodes	12 nodes	15 nodes	18 nodes
Not-Galen	43	42.27	41.06	39.12	36.70
GO	46.20	49.39	51.83	52.44	53.62
NCI	275	168.96	157.36	156.45	156.82
SNOMED	1,610.00	1,355.81	865.89	886.44	613.53
SNOMED-DUP-2	3,238.19	2,687.75	1,699.73	1,765.31	1,255.87
SNOMED-DUP-3	4,880.78	4,052.00	2,570.29	2,644.40	1,825.51
SNOMED-GALEN-GO	TIME OUT	TIME OUT	TIME OUT	TIME OUT	1,336.28

Ontology	21 nodes	25 nodes	28 nodes	32 nodes
Not-Galen	37.51	36.69	36.11	35.09
GO	51.89	56.76	39.70	50.80
NCI	155.19	154.75	161.41	160.12
SNOMED	529.30	441.74	442.81	383.01
SNOMED-DUP-2	1,064.44	887.19	893.96	755.38
SNOMED-DUP-3	1,571.43	1278.62	1286.50	1146.71
SNOMED-GALEN-GO	1,241.96	702.02	693.51	618.18

Table 5. Classification time (in seconds) of DistEL

$\mathcal{EL}^+$

- Concepts in $\mathcal{EL}^+$ are formed according to the grammar

$$C ::= A \mid \top \mid C \sqcap D \mid \exists r.C,$$

where A ranges over concept names, r over role names, and C, D over (possibly complex) concepts.

- An ontology in $\mathcal{EL}^+$ is a finite set of *general concept inclusions* $C \sqsubseteq D$ and *role inclusions* $r_1 \circ \cdots \circ r_n \sqsubseteq r$, where $r, r_1, \ldots, r_n$ are role names, $n \in \mathbb{Z}^+$.

Reasoning task: classification

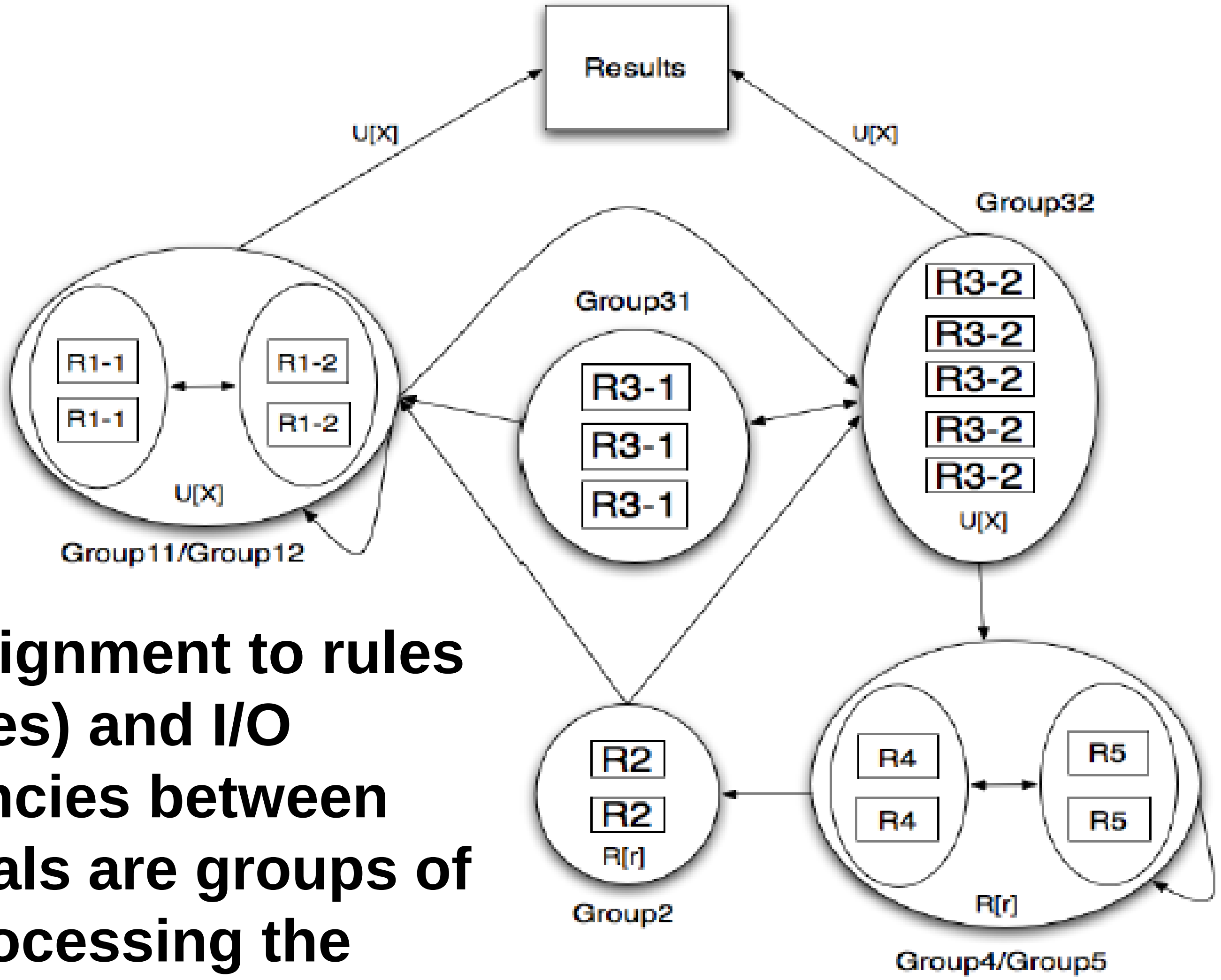
Compute all $A \sqsubseteq B$ for named classes A, B.

Completion-rule-based algorithm

Normal Form		Completion Rule
$A \sqsubseteq B$	**R1-1**	If $A \in S(X)$, $A \sqsubseteq B \in \mathcal{O}$, and $B \notin S(X)$ then $S(X) := S(X) \cup \{B\}$
$A_1 \sqcap \cdots \sqcap A_n \sqsubseteq B$	**R1-2**	If $A_1, \ldots, A_n \in S(X)$, $A_1 \sqcap \cdots \sqcap A_n \sqsubseteq B \in \mathcal{O}$, $B \notin S(X)$ then $S(X) := S(X) \cup \{B\}$
$A \sqsubseteq \exists r.B$	**R2**	If $A \in S(X)$, $A \sqsubseteq \exists r.B \in \mathcal{O}$, and $(X, B) \notin R(r)$ then $R(r) := R(r) \cup \{(X, B)\}$
$\exists r.A \sqsubseteq B$	**R3-1**	If $A \in S(Y)$, $\exists r.A \sqsubseteq B \in \mathcal{O}$ then $P = P \cup \{\exists r.Y \sqsubseteq B\}$
$\exists r.A \sqsubseteq B$	**R3-2**	If $(X, Y) \in R(r)$, $\exists r.Y \sqsubseteq B \in P$ and $B \notin S(X)$ then $S(X) := S(X) \cup \{B\}$
$r \sqsubseteq s$	**R4**	If $(X, Y) \in R(r)$, $r \sqsubseteq s \in \mathcal{O}$, and $(X, Y) \notin R(s)$ then $R(s) := R(s) \cup \{(X, Y)\}$
$r \circ s \sqsubseteq t$	**R5**	If $(X, Y) \in R(r)$, $(Y, Z) \in R(s)$, $r \circ s \sqsubseteq t \in \mathcal{O}$, $(X, Z) \notin R(t)$ then $R(t) := R(t) \cup \{(X, Z)\}$

Table : Axioms (in normal forms) and modified completion rules of CEL

Distribution



Node assignment to rules (rectangles) and I/O dependencies between rules. Ovals are groups of nodes processing the same rules.

Implementation/Optimizations

- Redis key-value store.
- Choice of keys and values.
- Data encoding (numerical identifiers)
- Selective data duplication
- Highly targeted communication (to relevant nodes only)
- Etc.

- Each node in our test cluster has two quad-core AMD Opteron 2300 MHz processors with 16GB RAM.

Future Work

- **Automated load balancing.**
- **Automated assignment of rules to nodes.**
- **Add ABox reasoning**
- **Other rulesets**

Related Work

Very little on distributed memory OWL reasoning.

Even less with convincing evaluations.

But see e.g.

- **Urbani et al on WebPIE and QueryPIE.**
- **Schlicht and Stuckenschmidt on distributed resolution for description logics.**

Conclusions

- This is the very first presentation of a distributed memory $\mathcal{EL}^+$ reasoner with convincing evaluation regarding parallelization and control of communication overhead.
- We seem to have significant scope for further optimizations.
- Generalizability of the architecture remains to be investigated.

Thanks!

Thanks!

Implementation-specific questions should best go to Raghava Mutharaju, mutharaju.2@wright.edu.

References

- Jacopo Urbani, Robert Piro, Frank van Harmelen, Henri Bal, Hybrid Reasoning on OWL RL. Semantic Web journal, to appear.
- Jacopo Urbani, Spyros Kotoulas, Jason Maassen, Frank van Harmelen, Henri E. Bal: WebPIE: A Web-scale Parallel Inference Engine using MapReduce. J. Web Sem. 10: 59-75 (2012)
- Anne Schlicht, Heiner Stuckenschmidt: Peer-to-Peer Reasoning for Interlinked Ontologies. Int. J. Semantic Computing 4(1): 27-58 (2010)
- Anne Schlicht, Heiner Stuckenschmidt: MapResolve. In: Proc. RR 2011: 294-299
- Raghava Mutharaju, Frederick Maier, Pascal Hitzler, A MapReduce Algorithm for EL+. In: Volker Haarslev, Davind Toman, Grant Weddell (eds.), In: Proc. DL2010, pp. 464-474.
- Kathrin Dentler, Ronald Cornet, Annette ten Teije, Nicolette de Keizer: Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web 2(2): 71-87 (2011)
- Franz Baader, Sebastian Brandt, Carsten Lutz: Pushing the EL Envelope. In: Proc. IJCAI 2005: 364-369