Evaluating Storage and Reasoning Systems

Mikalai Yatskevich

Information Systems Group

Oxford University Computing Laboratory, United Kingdom

mikalai.yatskevich@comlab.oc.ac.uk

Index

- Evaluation scenarios
- Evaluation descriptions
- Test data
- Tools
- Results
- Conclusion

Advanced reasoning system

- Description logic based system (DLBS)
- Standard reasoning services
  - Classification
  - Class satisfiability
  - Ontology satisfiability
  - Logical entailment

Existing evaluations

- Datasets
  - Synthetic generation
  - Hand crafted ontologies
  - Real-world ontologies
- Evaluations
  - KRSS benchmark
  - TANCS benchmark
  - Gardiner dataset

Evaluation criteria

- Interoperability
  - the capability of the software product to interact with one or more specified systems
  - a system must
    - conform to the standard input formats
    - be able to perform standard inference services
- Performance
  - the capability of the software to provide appropriate performance, relative to the amount of resources used, under stated conditions

Evaluation metrics

- Interoperability
  - Number of tests passed without parsing errors
  - Number of inference tests passed
- Performance
  - Loading time
  - Inference time

Class satisfiability evaluation

- Standard inference service that is widely used in ontology engineering
- The goal: to assess both DLBS's interoperability and performance
- Input
  - OWL ontology
  - One or several class IRIs
- Output
  - TRUE the evaluation outcome coincide with expected result
  - FALSE the evaluation outcome differ from expected outcome
  - ERROR indicates IO error
  - UNKNOWN indicates that the system is unable to compute inference in the given timeframe

Class satisfiability evaluation
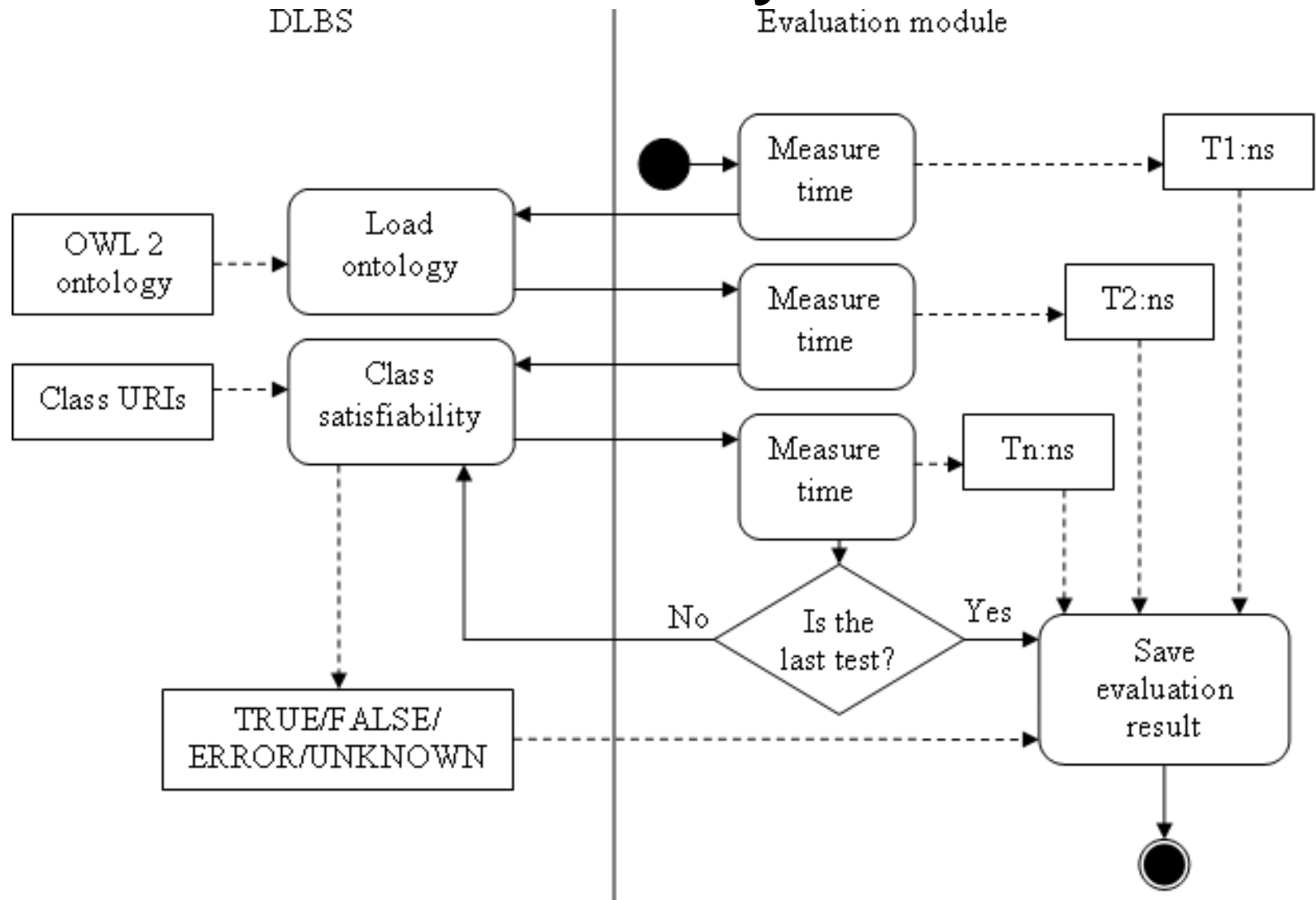
Ontology satisfiability evaluation

- Standard inference service typically carried out before performing any other reasoning task
- The goal: to assess both DLBS's interoperability and performance
- Input
  - OWL ontology
- Output
  - TRUE the evaluation outcome coincide with expected result
  - FALSE the evaluation outcome differ from expected outcome
  - ERROR indicates IO error
  - UNKNOWN indicates that the system is unable to compute inference in the given timeframe
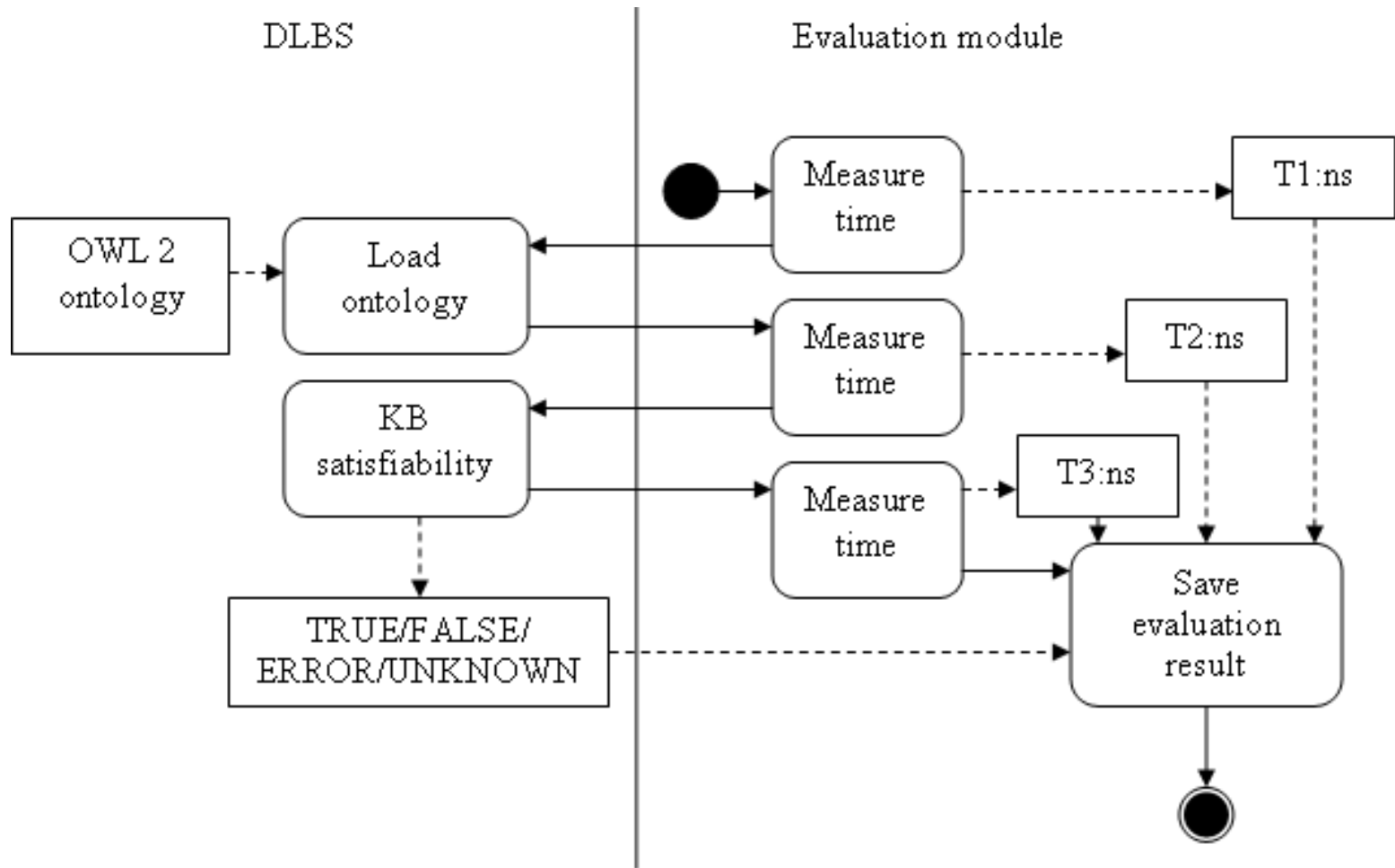
Ontology satisfiability evaluation

Classification evaluation

- Inference service that is typically carried out after testing ontology satisfiability and prior to performing any other reasoning task
- The goal: to assess both DLBS's interoperability and performance
- Input
  - OWL ontology
- Output
  - OWL ontology
  - ERROR indicates IO error
  - UNKNOWN indicates that the system is unable to compute inference in the given timeframe

Classification evaluation
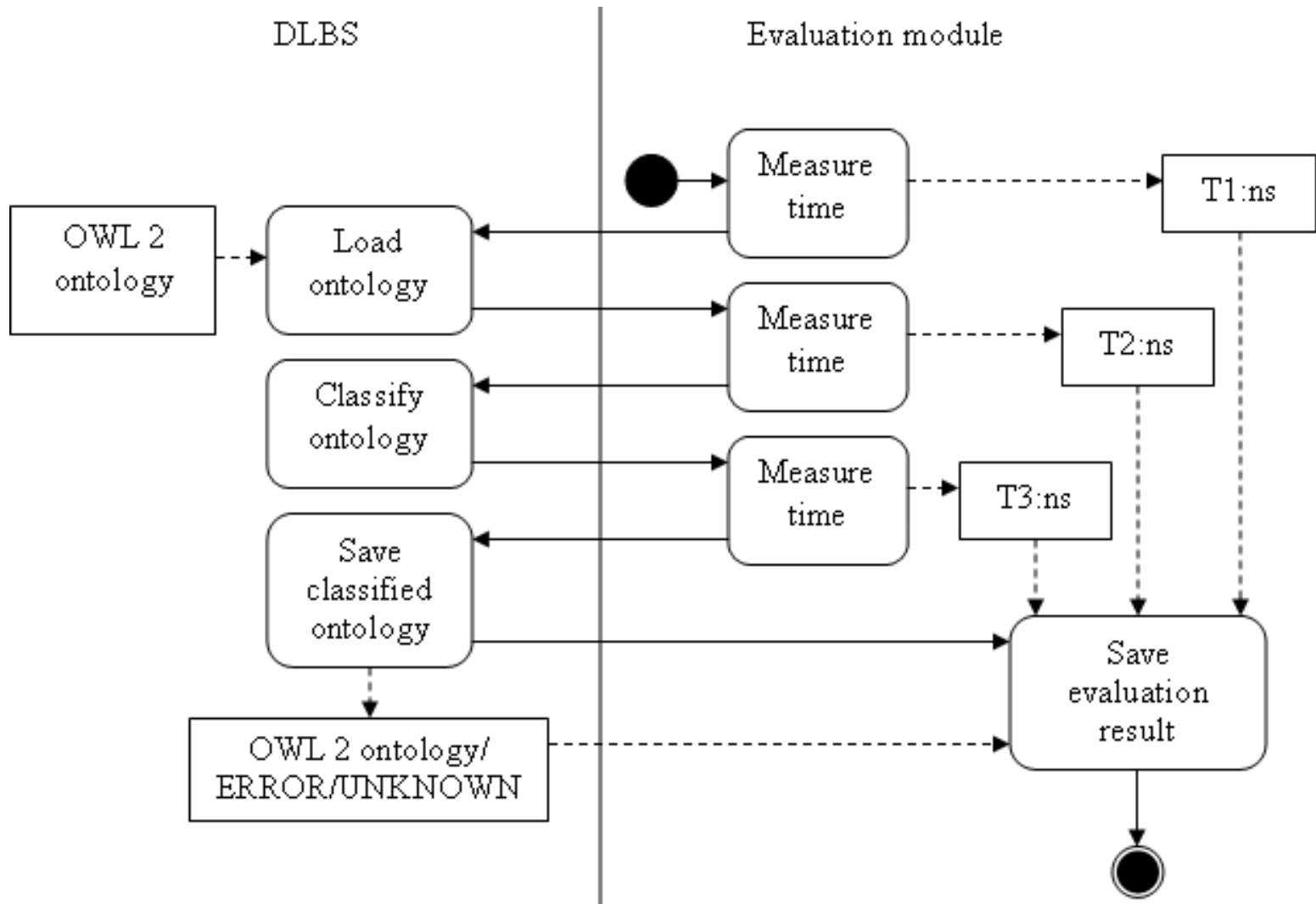
Logical entailment evaluation

- Standard inference service that is the basis for query answering
- The goal: to assess both DLBS's interoperability and performance
- Input
  - 2 OWL ontologies
- Output
  - TRUE the evaluation outcome coincide with expected result
  - FALSE the evaluation outcome differ from expected outcome
  - ERROR indicates IO error
  - UNKNOWN indicates that the system is unable to compute inference in the given timeframe

Logical entailment
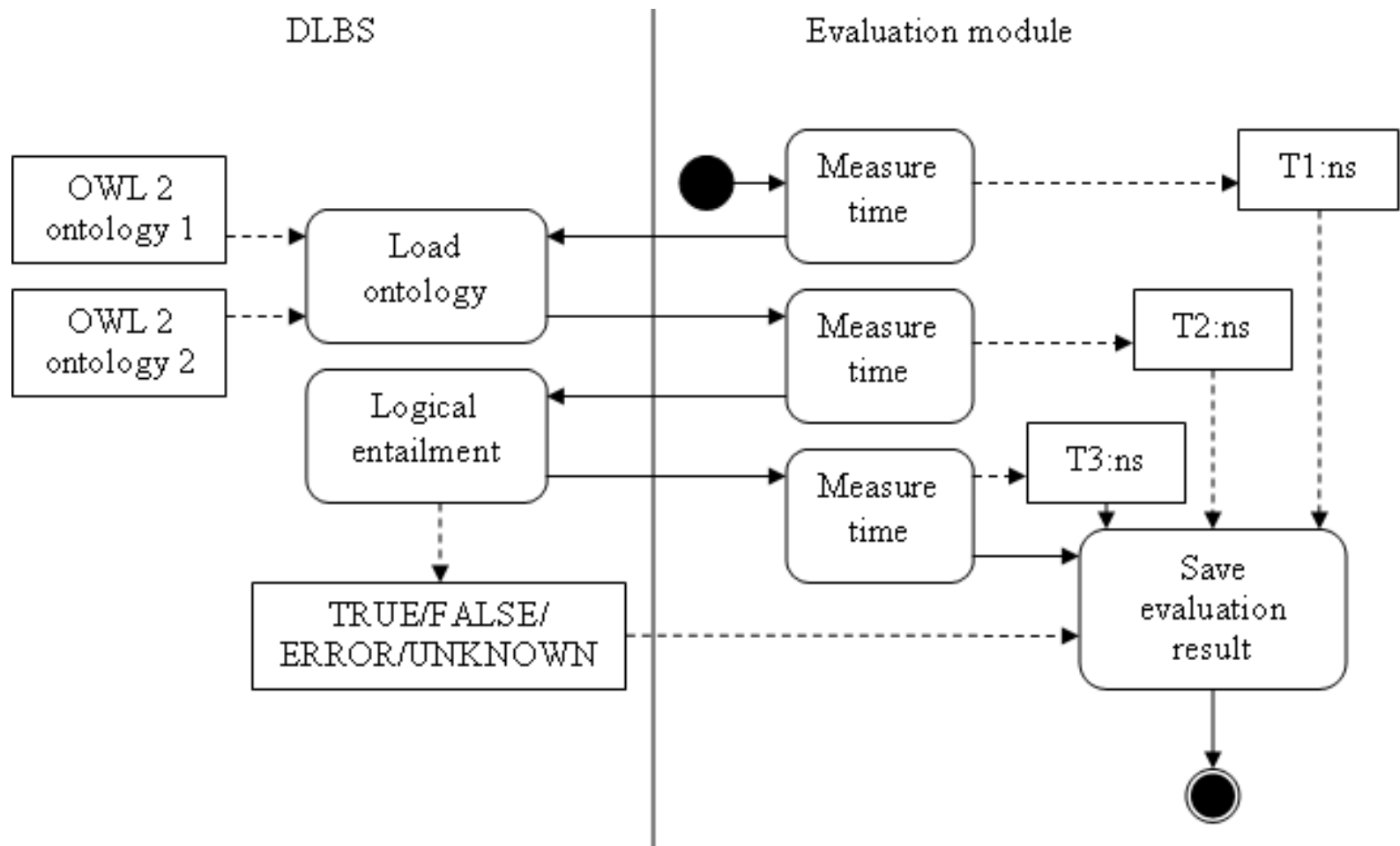
Storage and reasoning systems evaluation component

- SRS component is intended to evaluate the description logic based systems (DLBS)

  - Implementing OWL-API 3 de-facto standard for DLBS
  - Implementing SRS SEALS DLBS interface

- SRS supports test data in all syntactic formats supported by OWL-API 3

- SRS saves the evaluation results and interpretations in MathML 3 format

DLBS interface

- Java methods to be implemented by system developers
  - OWLOntology loadOntology(IRI iri)
  - boolean isSatisfiable(OWLOntology onto, OWLClass class)
  - boolean isSatisfiable(OWLOntology onto)
  - OWLOntology classifyOntology(OWLOntology onto)
  - URI saveOntology(OWLOntology onto, IRI iri)
  - boolean entails(OWLOntology onto1, OWLOntology onto2)

Testing Data

- The ontologies from the Gardiner evaluation suite.

  - Over 300 ontologies of varying expressivity and size.

- Various versions of the GALEN ontology

- Various ontologies that have been created in EU funded projects, such as SEMINTEC, VICODI and AEO

- 155 entailment tests from OWL 2 test cases repository

Evaluation setup

- 3 DLBSs
  - FaCT++ C++ implementation of FaCT OWL DL reasoner
  - HermiT Java based OWL DL reasoner utilizing novel hypertableau algorithms
  - Jcel Java based OWL 2 EL reasoner
- 2 AMD Athlon(tm) 64 X2 Dual Core Processor 4600+ machines with 2GB of main memory
  - DLBSs were allowed to allocate up to 1 GB

Evaluation results: Classification

	FaCT++	HermiT	jcel
ALT, ms	68		856
ART, ms	15320		2144
TRUE	160		16
FALSE	0		0
ERROR	47		4
UNKNOWN	3		0

Evaluation results: Class satisfiability

	FaCT++	HermiT	jcel
ALT, ms	1047	255	438
ART, ms	21376	517043	1113
TRUE	157	145	15
FALSE	1	0	0
ERROR	36	35	5
UNKNOWN	16	30	0

Evaluation results: Ontology satisfiability

	FaCT++	HermiT	jcel
ALT, ms	1315		708
ART, ms	25175		1878
TRUE	134		16
FALSE	0		0
ERROR	45		4
UNKNOWN	0		0

Evaluation results: Entailment

	FaCT++	HermiT
ALT, ms	14	33
ART, ms	1	20673
TRUE	46	119
FALSE	67	14
ERROR	34	9
UNKNOWN	0	3

Evaluation results: Non entailment

	FaCT++	HermiT
ALT, ms	47	92
ART, ms	5	127936
TRUE	7	7
FALSE	0	1
ERROR	3	1
UNKNOWN	0	1

Conclusion

- Errors:
  - datatypes not supported in the systems
  - syntax related : a system was unable to register a role or a concept
  - expressivity errors
- Execution time is dominated by small number of hard problems