Sixth International Competition on Computational Models of Argumentation

Iosif Apostolakis Andrei Popescu Johannes P. Wallner

Graz University of Technology

Der Wissenschaftsfonds.

Overview of ICCMA 2025

Goal

- assess state of the art in systems for argumentative reasoning
- collect and re-distribute a heterogeneous benchmark set

Outline

- Setup of ICCMA'25: tracks and subtracks
- Benchmark submissions and selection
- Solver submissions
- Computation and checking correctness
- Results and ranking

Overview of ICCMA 2025

Goal

- assess state of the art in systems for argumentative reasoning
- collect and re-distribute a heterogeneous benchmark set

Outline

- Setup of ICCMA'25: tracks and subtracks
- Benchmark submissions and selection
- Solver submissions
- Computation and checking correctness
- Results and ranking

Thanks

Thanks to all developers of solvers, all developers of benchmark suites, and previous organizers of ICCMA!
The computational results have been achieved using the Austrian Scientific Computing (ASC) infrastructure.

Semantics and Reasoning Tasks

Argumentation semantics:

- Complete (CO)
- Stable (ST)
- Preferred (PR)
- Semi-stable (SST)
- Ideal (ID)

Reasoning tasks:

- Credulous acceptance (DC)
- Skeptical acceptance (DS)
- Single extension (SE)

Discontinued semantics/tasks:

- enumeration (EE) since 2021
- counting (CE) since 2023
- grounded semantics (GR) since 2021
- stage semantics (STG) since this edition

Subtracks

Combining of a semantics and reasoning task is a subtrack.

- DC-CO, DC-ST, DC-SST, DC-ID
- DS-PR, DS-ST, DS-SST
- SE-PR, SE-ST, SE-SST, SE-ID

There are some well-known equivalences:

- DC-CO = DC-PR (and DC-AD for admissible sets)
- DC-ID = DS-ID

- Solvers can choose subtracks to participate
- Several subtracks constitute a track

Main Track

- exact solvers for abstract argumentation
- sequential and open-source, no portfolios
- No-limits track: same, but parallel, portfolio-based, and closed-source allowed
- subtracks: all (except DC-ID, since we have SE-ID)

Submitted Solvers

- AFCA: Obiedkov and Sertkaya
- ASSAT solver family: Liu, Dong, Ji, Ding, Luo, He, Xiong, Wu, and Ke
- Fudge: Thimm, Cerutti, and Vallati
- mu-toksia: Niskanen and Järvisalo
- reducto: Bengel, Sander, and Thimm
- Scallop: Lagniez, Lonca, and Mailly
- SMART: Hoffmann, Kuhlmann, and Thimm

Heuristics Track

- inexact solvers for abstract argumentation
- no requirement of returning correct solutions
- incorrect solutions discarded
- subtracks: all for DC and DS

Submitted Solvers

- ARIPOTER: Cibier, Delobelle, Mailly, and Rossit
- Fargo-Timelimited-FLiGS: Cibier and Mailly
- FARGO-LIMITED: Thimm
- FastAFGCN: Malmqvist
- FastLiGS: Cibier and Mailly
- HARPER++: Thimm
- Heuback: Malmqvist
- Scallop: Lagniez, Lonca, and Mailly

Dynamics Track

- sequence of queries of changing argumentation framework
- requirement to implement the IPAFAIR API
- subtracks: DC-CO, DC-ST, DS-ST, DS-PR

Submitted Solvers

- mu-toksia: Niskanen and Järvisalo
- Scallop: Lagniez, Lonca, and Mailly

Assumption-based Argumentation (ABA) Track

- exact solvers for flat ABA frameworks (“logic programming fragment”)
- subtracks: DC-CO, DC-ST, DS-PR, DS-ST, SE-PR, SE-ST

Submitted Solvers

- 100BA: Niskanen, Rankooh, Lehtonen, and Järvisalo
- AcbAr: Lehtonen, Rapberger, and Ulbricht
- ASPforABA: Lehtonen and Järvisalo
- MS-DIS: Diller and Gorczyca
- Scallop: Lagniez, Lonca, and Mailly

Resouce Limits for Main, Dynamic, and ABA

- Time limit: 1200 seconds (wall time for No-limits track, CPU time for other tracks)
- Memory: 32 GB

PAR-2 Scoring for Main, Dynamic, and ABA

- score of instance: $2 \cdot 1200$ if limit reached
- score of subtrack: sum over all instances
- ranking: from lowest to highest

Heuristics Track

- Time limit: 60 seconds CPU
- Memory: 32 GB
- ranking: number of correctly solved instances (CPU time as tie-braker)

Fixed SAT Solver Comparison

- Solvers implement the IPASIR interface for (incremental) SAT solving
- Organizers fix the SAT solver
- this edition: CaDiCaL version 2.1.3

Participating Solvers

- Fudge: Thimm, Cerutti, and Vallati
- mu-toksia: Niskanen and Järvisalo
- reducto: Bengel, Sander, and Thimm
- Scallop: Lagniez, Lonca, and Mailly

AF Input

```
p af 5
# this is a comment
1 2
2 4
4 5
5 4
5 5
```

ABA Input

```
p aba 8
# this is a comment
a 1
a 2
a 3
c 1 6
c 2 7
c 3 8
r 4 5 1
r 5
r 6 2 3
```

Output for DC/DS

YES or NO

Output for SE

w 1 2 4

... for extension (or set) $\{1, 2, 4\}$

Witnesses (only Main/No-Limits Track)

- DC and YES answer:
  - DC-CO: admissible set containing the queried argument.
  - Other cases witness extension of the semantics with argument.
- DS and NO answer:
  - DS-PR: preferred extension excluding argument or admissible set attacking argument.
  - Other cases: extension of the semantics excluding the argument.

For instance, for DC-CO:

```
YES
w 1 2 4
```

Witnesses were verified with a similar procedure as last edition

- SAT-based checking (using standard encodings)
- when UNSAT is required: SAT proof and check
- for SE-ID: check if solvers report the same
- For heuristics track: results from Main Track; reduced “unknowns” by higher timeout (lead to 66 unknowns)

- We reported bugs to solver developers
- Some minor issues: we let the developers bugfix the solvers

Benchmark Suite for AFs (Main/No-Limit/Heuristics/Dynamics)

from ICCMA'17, ICCMA'19, and ICCMA'23:

- GroundedGenerator, SCCGenerator, StableGenerator,
- ABA2AF,
- AdmBuster,
- a generator for AF families called Barabasi-Albert, Erdos-Renyi and Watts-Strogatz,
- PIanning2AF,
- Sembuster,
- Traffic,
- AFGen,
- Datalog, and
- crusti_g2io

and two new ones:

- KWT Benchmark Generator by Kuhlmann and Thimm
- Subsampling Generator by Malmqvist

Sampling & Instance Selection

- up to 20 AFs from existing benchmarks
- up to 30 AFs from new benchmarks
- excluding AFs with less than 100 arguments
- query argument: uniformly at random, except self-attacking and arguments without incoming attacks

Result: 322 AFs

Benchmark Suite for ABAs

- RandomABAGenerator from previous edition
- ABCgen: by Niskanen, Rankooh, Lehtonen, and Järvisalo
  - crafted to be “hard”
- Benchmarks from the wild: by Odekerken
  - generator inspired from real-world application
  - originally intended for different reasoning tasks (not in ICCMA)
  - for ICCMA reasoning tasks: computationally simple

Sampling & Instance Selection

- 80 ABAs from previous generator
- 120 each for new generators
- query: uniformly at random from atoms
  - derivable from non-self contrary assumptions
  - not derivable only by assumptions without contraries

Result: 320 ABAs

DC-CO (average PAR-2 score) - Main Track

1. mu-toksia (88.22279)
2. Fudge (89.44603)
3. Scallop (93.55768)
4. reducto (109.41719)
5. AFCA (195.75326)
6. SMART (510.70802)

DC-CO (average PAR-2 score) - Main Track

1. mu-toksia (88.22279)
2. Fudge (89.44603)
3. Scallop (93.55768)
4. reducto (109.41719)
5. AFCA (195.75326)
6. SMART (510.70802)

DC-ST (average PAR-2 score) - Main Track

1. Fudge (72.93384)
2. mu-toksia (75.42434)
3. Scallop (83.30070)
4. reducto (108.32354)

DC-CO (average PAR-2 score) - Main Track

1. mu-toksia (88.22279)
2. Fudge (89.44603)
3. Scallop (93.55768)
4. reducto (109.41719)
5. AFCA (195.75326)
6. SMART (510.70802)

DC-ST (average PAR-2 score) - Main Track

1. Fudge (72.93384)
2. mu-toksia (75.42434)
3. Scallop (83.30070)
4. reducto (108.32354)

DC-SST (average PAR-2 score) - Main Track

1. mu-toksia (136.6354)
2. Fudge (138.6929)
3. Scallop (196.5304)

DS-PR (average PAR-2 score) - Main Track

1. mu-toksia (127.1715)
2. reducto (127.9055)
3. Fudge (137.8313)
4. Scallop (178.5025)

DS-PR (average PAR-2 score) - Main Track

1. mu-toksia (127.1715)
2. reducto (127.9055)
3. Fudge (137.8313)
4. Scallop (178.5025)

DS-ST (average PAR-2 score) - Main Track

1. Fudge (103.1939)
2. mu-toksia (106.2097)
3. Scallop (107.4829)
4. reducto (139.1873)

DS-PR (average PAR-2 score) - Main Track

1. mu-toksia (127.1715)
2. reducto (127.9055)
3. Fudge (137.8313)
4. Scallop (178.5025)

DS-ST (average PAR-2 score) - Main Track

1. Fudge (103.1939)
2. mu-toksia (106.2097)
3. Scallop (107.4829)
4. reducto (139.1873)

DS-SST (average PAR-2 score) - Main Track

1. Scallop (172.9995)
2. mu-toksia (190.4484)
3. Fudge (198.7965)

SE-ID (average PAR-2 score) - Main Track

1. fudge (200.1302)
2. mutoksia (200.2320)
3. scallop (416.1002)

SE-PR (average PAR-2 score) - Main Track

1. Fudge (97.6889)
2. mu-toksia (108.7563)
3. Scallop (116.2437)
4. reducto (153.1086)
5. ASSAT-amsat (180.0773)
6. ASSAT-llm (191.7278)
7. ASSAT-inpriority (222.0546)
8. ASSAT-cadical (222.9127)

SE-PR (average PAR-2 score) - Main Track

1. Fudge (97.6889)
2. mu-toksia (108.7563)
3. Scallop (116.2437)
4. reducto (153.1086)
5. ASSAT-amsat (180.0773)
6. ASSAT-llm (191.7278)
7. ASSAT-inpriority (222.0546)
8. ASSAT-cadical (222.9127)

SE-ST and SE-SST

... postponed (as well as No-Limits)

DC-CO (correct instances) - Heuristics

1. Scallop (304)
2. Fargo-Timelimited-FLiGS (301)
3. fargolimited (282)
4. FastLiGS (246)
5. FastAFGCN (226)
6. HARPER++ (225)
7. ARIPOTER (222)
8. Heuback (190)

DC-ST (correct instances) - Heuristics

1. Scallop (310)
2. FastLiGS (268)
2. Fargo-Timelimited-FLiGS (268)
4. fargolimited (251)
5. FastAFGCN (241)
6. ARIPOTER (226)
7. HARPER++ (178)
8. Heuback (152)

DC-SST (correct instances) - Heuristics

1. Scallop (303)
2. Fargo-Timelimited-FLiGS (296)
3. fargolimited (276)
4. FastLiGS (245)
5. ARIPOTER (232)
6. FastAFGCN (220)
7. HARPER++ (214)
8. Heuback (184)

DC-ID (correct instances) - Heuristics

1. Fargo-Timelimited-FLiGS (301)
2. Heuback (299)
2. HARPER++ (299)
4. FastAFGCN (295)
5. ARIPOTER (293)
6. fargolimited (284)
7. FastLiGS (277)
8. Scallop (252)

DS-PR (correct instances) - Heuristics

1. Fargo-Timelimited-FLiGS (306)
2. ARIPOTER (304)
2. HARPER++ (304)
4. Heuback (303)
5. FastAFGCN (298)
6. Scallop (291)
7. fargolimited (289)
8. FastLiGS (284)

DS-ST (correct instances) - Heuristics

1. Scallop (307)
2. FastLiGS (252)
3. Fargo-Timelimited-FLiGS (224)
4. Heuback (219)
5. fargolimited (191)
6. HARPER++ (189)
7. FastAFGCN (168)
8. ARIPOTER (167)

DS-SST (correct instances) - Heuristics

1. Fargo-Timelimited-FLiGS (296)
2. HARPER++ (295)
2. Heuback (295)
4. ARIPOTER (289)
4. FastAFGCN (289)
6. fargolimited (285)
7. FastLiGS (281)
7. Scallop (281)

(average PAR-2 score) - Dynamics Track

	DC-CO	DC-ST	DS-ST	DS-PR*
Scallop	354.7872	263.1717	238.6814	229.6096
mu-toksia	450.0123	427.5895	510.9314	580.9538

* Remark: solvers disagree on one instance in DS-PR.

DC-CO (average PAR-2 score) - ABA Track

1.	ASPforABA (293.9361)
2.	100BA-ufs (364.7527)
3.	100BA-acyc (465.8241)
4.	100BA-ve (479.1673)
5.	Scallop (652.2234)
6.	AcbAr (819.2310)
7.	MS-DIS (1242.0169)

DC-CO (average PAR-2 score) - ABA Track

1. ASPforABA (293.9361)
2. 100BA-ufs (364.7527)
3. 100BA-acyc (465.8241)
4. 100BA-ve (479.1673)
5. Scallop (652.2234)
6. AcbAr (819.2310)
7. MS-DIS (1242.0169)

DC-ST (average PAR-2 score) - ABA Track

1. ASPforABA (248.3296)
2. 100BA-ufs (324.9272)
3. 100BA-acyc (415.7108)
4. 100BA-ve (441.6975)
5. Scallop (600.4252)
6. AcbAr (683.9565)
7. MS-DIS (2370.3427)

DS-PR (average PAR-2 score) - ABA Track

1.	ASPforABA (343.5808836)
2.	100BA-ufs (499.2441503)
3.	100BA-ve (558.0177546)
4.	Scallop (670.3195038)
5.	100BA-acyc (701.7749629)
6.	AcbAr (936.8117998)

DS-PR (average PAR-2 score) - ABA Track

1.	ASPforABA (343.5808836)
2.	100BA-ufs (499.2441503)
3.	100BA-ve (558.0177546)
4.	Scallop (670.3195038)
5.	100BA-acyc (701.7749629)
6.	AcbAr (936.8117998)

DS-ST (average PAR-2 score) - ABA Track

1.	ASPforABA (225.1797)
2.	100BA-ufs (358.3513)
3.	100BA-ve (417.2195)
4.	100BA-acyc (441.2361)
5.	Scallop (551.9113)
6.	AcbAr (750.5788)

SE-PR (average PAR-2 score) - ABA Track

1.	ASPforABA (190.3292)
2.	100BA-ufs (324.2567)
3.	100BA-ve (431.2813)
4.	100BA-acyc (465.8177)
5.	Scallop (578.5660)
6.	AcbAr (805.8363)

SE-PR (average PAR-2 score) - ABA Track

1.	ASPforABA (190.3292)
2.	100BA-ufs (324.2567)
3.	100BA-ve (431.2813)
4.	100BA-acyc (465.8177)
5.	Scallop (578.5660)
6.	AcbAr (805.8363)

SE-ST (average PAR-2 score) - ABA Track

1.	ASPforABA (162.0906)
2.	100BA-ufs (247.5586)
3.	100BA-ve (371.4530)
4.	100BA-acyc (395.9575)
5.	Scallop (532.8720)
6.	AcbAr (654.7186)

Summary

- Several new submissions, solvers and benchmarks
- New: Fixed SAT Solver Comparison
- Semantics list slightly shortened
- `argumentationcompetition.org/2025/index.html`

ABA Track

- While ASPforABA won the tracks, instances unsolved by ASPforABA and solved by others exist
- Novel Benchmarks: hard crafted ones and real-world ones

More Detailed Results

- Will be published and announced
- “Missing” results will be announced
- Descriptions will be published