

# Summary Report of the Second International Competition on Computational Models of Argumentation

*Sarah A. Gaggl, Thomas Linsbichler,  
Marco Maratea, Stefan Woltran*

■ *We give an overview of the design and results of the Second International Competition on Computational Models of Argumentation (ICCMA-17). Following the first meeting in 2015, the competition evaluates the performance of submitted solvers on computational problems within abstract argumentation. In addition to the four original semantics, ICCMA-17 includes three additional prominent semantics. Moreover, a dedicated call for benchmarks allowed for the introduction of a sophisticated instance selection process.*

Argumentation is a major topic in the study of artificial intelligence (Bench-Capon and Dunne 2007; Atkinson et al. 2017). In particular, the problem of solving certain reasoning tasks of Dung’s abstract argumentation frameworks (Dung 1995) is central to many advanced argumentation systems. The fact that problems to be solved are mostly intractable requires efficient algorithms and solvers, which are to be evaluated on meaningful benchmarks. Another unique feature of abstract argumentation is the fact that solvers are expected to handle different semantics. This feature makes the design of competitions quite different from other comparable events, for instance in the field of propositional logic (SAT) or answer-set programming (ASP).

In this report, we briefly present the design and results of the Second International Competition on Computational Models of Argumentation (ICCMA-17),<sup>1</sup> which was jointly organized by TU Dresden (Germany), TU Wien (Austria), and the University of Genoa (Italy), in affiliation with the 2017 International Workshop on Theory and Applications of Formal Argumentation (TAFAs-17). ICCMA-17 was conducted in the first half of 2017 and comes two years after the first competition, ICCMA-15<sup>2</sup> (Thimm et al. 2016).

The general goal of this competition is to consolidate and strengthen the ICCMA series, which in its first meeting had very good outcomes in some respects, for example, in terms of the number of submitted solvers (18). The 2017 competition maintained some of the design choices previously made, such as the I/O formats and the basic reasoning problems. With a slight modification to the first meeting, the competition was organized into tasks and tracks, where a task is a reasoning problem under a particular semantics, and a track collects different tasks over a semantics. ICCMA-17 also introduced several novelties: (1) a new scoring scheme for better reflecting the solvers' behavior, (2) three new semantics, namely semistable, stage, and ideal semantics, (3) a special Dung's Triathlon track, where solvers were required to deal with different problems simultaneously, with the goal of testing the solvers' capability for exploiting interrelationships between semantics, and (4) a call for benchmarks to enrich the suite of instances for the competition, followed by a novel instance selection stage.

## Background and Format

An abstract argumentation framework (AF, for short) (Dung 1995) is a tuple  $\mathcal{F} = (A, \rightarrow)$  where  $A$  is a set of arguments and  $\rightarrow \subseteq A \times A$  is the attack relation. Semantics are used to determine sets of jointly acceptable arguments by mapping each AF to a set of extensions  $\sigma(\mathcal{F}) \subseteq 2^A$  (see Baroni, Caminada, and Giacomin [2011] for an overview). The main underlying concepts of semantics are conflict-freeness and admissibility. The semantics considered in the competition were grounded, complete, preferred, stable (Dung 1995), semistable (Caminada, Carnielli, and Dunne 2012), stage (Verheij 1996), and ideal (Dung, Mancarella, and Toni 2007), the last three for the first time in the ICCMA series.

Following ICCMA-15, we considered four reasoning problems: skeptical and credulous acceptance; and computing a single extension and all extensions. Under the semantics considered, the complexity of these problems ranges from polynomial time to intractability in the second level of the polynomial hierarchy (Dunne and Wooldridge 2009; Dvorak 2012).

The competition featured seven main tracks, one

for each semantic category. Each of these tracks comprised four (two, respectively, for grounded and ideal semantics, given they are single-status) tasks, one for each reasoning problem. The reasoning problems combined with the semantics yielded a total of 24 tasks.

A special eighth track, the Dung's Triathlon, was conducted to enumerate three types of extensions simultaneously: grounded, stable, and preferred.

## Participants

Sixteen solvers participated in the competition, 10 of which were new entries compared to the 2015 competition. The solvers originated from 14 different teams from Austria, China, Finland, France, Germany, Italy, Jordan, and the UK. Each solver could compete in an arbitrary set of tasks. If a solver supported all tasks of a track, it also participated in the track. This strategy resulted in each task featuring at least nine participating solvers, and eight solvers participating in all tracks. The solvers participating in this second event were based on a wide variety of solving approaches, ranging from direct approaches to (different forms of) reductions to SAT, ASP, CSP, and circumscription.

All solver submissions were accompanied by a system description and the full source code, so as to ensure maximal transparency and accessibility to the community.

## Benchmarks and Selection

For the first time, ICCMA took advantage of a dedicated call for benchmarks, which is customary in other competitions. We received six submissions, among them both AF generators and concrete sets of AFs. The concrete AFs included collections of (1) AFs instantiated from assumption-based argumentation, (2) AFs translated from planning problems, and (3) AFs obtained from traffic network graphs. The submitted generators produced AFs crafted to be challenging for (4) strong admissibility and (5) semistable semantics and (6) AFs derived from well-known graph classes from the literature (Barabasi-Albert, Erdős-Rényi, and Watts-Strogatz). With the generators, we produced instances that aimed to cover a possibly broad range of difficulty.

Together with the generators from ICCMA-15, GroundedGenerator, SccGenerator, and StableGenerator (Thimm and Villata 2017), these sets contributed to the benchmark suite of ICCMA-17, for a total of 3990 instances in 11 domains. The benchmark suite included a heterogeneous set of benchmarks, that is, random, crafted, and application oriented. Starting from this suite, a benchmark selection process was applied to select the instances to be run in the competition. Following related competitions, such as the SAT and ASP competitions (Belov et al. 2014; Gebser,

Track	Solver	Points
Complete	pyglaf	1229/1400
Preferred	ArgSemSAT	1146/1400
Stable	pyglaf	1183/1400
Semistable	argmat-sat	1164/1400
Stage	argmat-sat	1065/1400
Grounded	CoQuiAAS	695/700
Ideal	pyglaf	585/700
Dung's Triathlon	argmat-dvisat	276/350

Table 1. Award Winners.

Maratea, and Ricca 2017) — but for the first time in the ICCMA competition — we selected instances based on their expected hardness, so as to have a benchmark suite covering a wide variety of expected difficulties. Given the high number of tasks and tracks in the competition, we organized the tasks already evaluated in ICCMA-15 into three groups, based on the complexity of the tasks. For each of these groups, we classified all instances into five hardness categories, from “very easy” to “too hard,” according to the performance of three of the best solvers from ICCMA-15 in a representative task within the group. The selection of these solvers ensured that they implemented different solving approaches in order to avoid biased results. For the tasks related to the newly introduced semantics and for Dung's Triathlon, we considered the classification obtained by the group containing the tasks of highest complexity. Then, after the classification, we selected 350 instances per group distributed over the hardness categories, ensuring also that instances were distributed over domains.

Finally, for the acceptance tasks, and considering that the number of instances had to be constant among tasks, we selected only one argument for each instance, with the exception that we dropped the “very easy” instances for acceptance tasks and selected two arguments to be queried for the “too hard” instances.

## Results

For each track, the score of a solver was obtained by the sum of scores over all tasks of the track, each of them obtained by the sum of points over all instances. For each instance, a solver received 1 point if it delivered the correct result within 600 (respectively 1800 for Dung's Triathlon) seconds of CPU

time, 5 points if it delivered an incorrect result, and 0 points otherwise. This represents a change from previously: in ICCMA-15 incorrect results were assigned 0 points, while in ICCMA-17, we assigned a negative reward. This change was applied to put the focus on correctness and to prevent solvers from guessing answers. We think in this way the final score better reflects the solvers' behavior. The correctness of results was verified by comparing the results to reference solutions by ASPARTIX, a reliable solver from ICCMA-15 (see Egly, Gaggl, and Woltran [2010]), by using dedicated ASP encodings for checking single extensions, and by comparing solutions between solvers.

Overall, the winner of a track was the solver that got the highest score. Ties were broken by the total time it took the solver to return correct results.

Table 1 lists the winner for each track, showing the semantics in the first column, the winner in the second column, and the number of points achieved by the winner in the third column. The exception is the last row, which contains the winner of the special Dung's Triathlon track. The list of winners reflects the diversity of the solving approaches in the competition and ultimately confirms the usefulness of having such a wide variety of approaches. Indeed, the five winning solvers implement concepts that take advantage of SAT (under different integration schemas), circumscriptions, and CSP techniques.<sup>3</sup>

## Conclusions and Outlook

The fact that two-thirds of tracks were won by solvers newly introduced at ICCMA-17 shows that the field of computational models of argumentation is not only vibrant but also highly amenable to further improvements and innovation. Moreover, pyglaf (the winner of three tracks) used a novel approach

based on reduction to circumscription, which indicates that even more variety of solving techniques can be fruitful for the development of the field.

While we think that future editions of ICCMA should stick to a guided instance selection process as described in this report, the community should aim for benchmarks from real-world domains to be included in future benchmark suites. On the technical side, changing the output format for enumeration tasks could be beneficial for the verification of large solutions.

The next competition will be conducted in 2019.<sup>4</sup>

### Notes

1. [argumentationcompetition.org/2017](http://argumentationcompetition.org/2017).
2. [argumentationcompetition.org/2015](http://argumentationcompetition.org/2015).
3. For detailed results, see [argumentationcompetition.org/2017/results.html](http://argumentationcompetition.org/2017/results.html).
4. For more information, see [argumentationcompetition.org](http://argumentationcompetition.org).

### Acknowledgments

This work has been supported by FWF (project I2854).

### References

Atkinson, K.; Baroni, P.; Giacomin, M.; Hunter, A.; Prakken, H.; Reed, C.; Simari, G. R.; Thimm, M.; and Villata, S. 2017. Towards Artificial Argumentation. *AI Magazine* 38(3): 25–36. doi.org/10.1609/aimag.v38i3.2704.

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An Introduction to Argumentation Semantics. *The Knowledge Engineering Review* 26(4): 365–410. doi.org/10.1017/S0269888911000166.

Belov, A.; Diepold, D.; Heule, M. J. H.; and Jarvisalo, M, eds. 2014. *Proceedings of SAT Competition 2014: Solver and Benchmark Descriptions*. Department of Computer Science Series of Publications B. Report B-2014-2. Helsinki, Finland: University of Helsinki.

Bench-Capon, T. J. M., and Dunne, P. E. 2007. Argumentation in Artificial Intelligence. *Artificial Intelligence* 171(10–15): 619–41. doi.org/10.1016/j.artint.2007.05.001.

Caminada, M.; Carnielli, W.; and Dunne, P. E. 2012. Semi-Stable Semantics. *Journal of Logic and Computation* 22(5): 1207–54. doi.org/10.1093/logcom/exr033.

Dung, P. M. 1995. On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and N-Person Games. *Artificial Intelligence* 77(2): 321–58. doi.org/10.1016/0004-3702(94)00041-X.

Dung, P. M.; Mancarella, P.; and Toni, F. 2007. Computing Ideal Sceptical Argumentation. *Artificial Intelligence* 171(10–15): 642–74. doi.org/10.1016/j.artint.2007.05.003.

Dunne, P. E., and Wooldridge, M. 2009. Complexity of Abstract Argumentation. In *Argumentation in Artificial Intelligence*, edited by I. Rahwan and G. Simari, 85–104. Berlin: Springer. doi.org/10.1007/978-0-387-98197-0\_5.

Dvorak, W. 2012. Computational Aspects of Abstract Argumentation. PhD dissertation, Institute of Information Systems, Technischen Universität Wien, Wien, Österreich.

Egly, U.; Gaggl, S. A.; and Woltran, S. 2010. Answer-Set Programming Encodings for Argumentation Frameworks. *Argument and Computation* 1(2): 147–77. doi.org/10.1080/19462166.2010.486479.

Gebser, M.; Maratea, M.; and Ricca, F. 2017. The Sixth Answer Set Programming Competition. *Journal of Artificial Intelligence Research* 60:41–95.

Thimm, M., and Villata, S. 2017. The First International Competition on Computational Models of Argumentation: Results and Analysis. *Artificial Intelligence* 252:267–94. doi.org/10.1016/j.artint.2017.08.006.

Thimm, M.; Villata, S.; Cerutti, F.; Oren, N.; Strass, H.; and Vallati, M. 2016. Summary Report of the First International Competition on Computational Models of Argumentation. *AI Magazine* 37(1): 102–4. doi.org/10.1609/aimag.v37i1.2640.

Verheij, B. 1996. Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages. In *Proceedings of the Eighth Dutch Conference on Artificial Intelligence*, edited by J. J. Meyer and L. C. van der Gaag, 357–68. Utrecht, the Netherlands: Utrecht University.

**Sarah A. Gaggl** is a postdoctoral research assistant at the Computational Logic Group at the TU Dresden. She received her PhD in computer science at TU Wien in 2013. Her main research interests are in abstract argumentation, answer set programming, and knowledge representation.

**Thomas Linsbichler** is a postdoctoral researcher at the Institute of Information Systems of TU Wien, Austria. His main research interests are in knowledge representation and reasoning, argumentation, and algorithms.

**Marco Maratea** is an associate professor in computer engineering at University of Genoa, Italy. In the fall of 2015, 2016, and 2017, he was a university lecturer at the Institute for Information Systems of the Faculty of Informatics at the Vienna University of Technology. His research interests include artificial intelligence, logic programming, and knowledge representation and reasoning.

**Stefan Woltran** is professor of foundations of artificial intelligence at Vienna University of Technology. His research focuses on problems in the area of knowledge representation and reasoning, argumentation, complexity analysis in artificial intelligence, and logic programming. In the winter term 2013, he held a deputy professorship at Leipzig University. In 2013, he also received the prestigious START award from the Austrian Science Fund (FWF).