

The Third (Open) Answer Set Programming Competition

F. Calimeri, G. Ianni, F. Ricca

Dipartimento di Matematica, Università della Calabria

11th International Conference on Logic Programming and
Nonmonotonic Reasoning

presented by G. Ianni, F. Ricca

Outline

- ① The Third ASP Competition
- ② System Track
- ③ Model and Solve Track
- ④ Comparisons

Outline

- 1 The Third ASP Competition
- 2 System Track
- 3 Model and Solve Track
- 4 Comparisons

The Third ASP Competition

An event gaining momentum and maturity

- Two tracks: the System Competition Track and the Model and Solve Competition Track
- Host institution: the Dipartimento di Matematica at Università della Calabria (one of the proud homes of the DLV System)

The Third ASP Competition

An event gaining momentum and maturity

- Two tracks: the System Competition Track and the Model and Solve Competition Track
- Host institution: the Dipartimento di Matematica at Università della Calabria (one of the proud homes of the DLV System)
- **System track:** for “pure” ASP systems
- **Model and Solve track:** open (ASP systems, CSP systems, PDDL based, FO(ID)...))

The Third ASP Competition

An event gaining momentum and maturity

- Two tracks: the System Competition Track and the Model and Solve Competition Track
- Host institution: the Dipartimento di Matematica at Università della Calabria (one of the proud homes of the DLV System)
- **System track:** for “pure” ASP systems
- **Model and Solve track:** open (ASP systems, CSP systems, PDDL based, FO(ID)...)
- System Track: 11 Participants + 7 Non-competing
- M&S Track: 6 Teams
- The host institution stood neutral

Benchmark Selection - I

- 35 problems (of which 18 for the System track only):

Benchmark Selection - I

- 35 problems (of which 18 for the System track only):

Field of provenance

- **Database/Information extraction:** Grammar-BasedInformationExtraction, Weight-AssignmentTree
- **Planning/Routing:** AirportPickup, HanoiTower, Sokoban, FastFood, Solitaire, KnightTour, Hydraulic Planning
- **Classic graph problems:** Reachability, MaximalClique, GraphColouring, CrossingMinimization, StableMarriage
- **Biology/Academy:** PartnerUnits, MinimalDiagnosis, ReverseFolding
- **Scheduling/Packing:** DisjunctiveScheduling, IncrementalScheduling, Packing
- **Puzzle inspired:** Tangram, MagicSquareSets, MazeGeneration, Numberlink, GeneralizedSlitherlink, Tomography

Benchmark Selection - II

Difficulty level

- Some problems (20) came from the 2nd Competition: harder instances and/or harder variant introduced whenever necessary;

Benchmark Selection - II

Difficulty level

- Some problems (20) came from the 2nd Competition: harder instances and/or harder variant introduced whenever necessary;

Polynomial, NP, Beyond-NP (Σ_2^P), Optimization (Model Track only):

Complexity/Type

- (Pure) Polynomial (5 problems). Natural declarative encoding is polynomial: mostly measures efficiency of grounders;
- (Known) Polynomial (2 problems). Natural declarative encoding includes guessing. Measures 'convergence' capabilities of systems (StableMarriage, PartnerUnitsPolynomial)
- NP (19 problems): the "core" class.
- Beyond-NP/Optimization (2 + 6 problems).

Scoring

Each participant is awarded the sum of scores per each benchmark domain, naturally weighing NP (>50%) more than P (20-25%) and Beyond-NP/Opt (20-25%).

Scoring

Each participant is awarded the sum of scores per each benchmark domain, naturally weighing NP (>50%) more than P (20-25%) and Beyond-NP/Opt (20-25%).

Instance Quota

$$S_{solve}(P) = \frac{N_S}{N}$$

N_S = number of solved instances for problem P , N = total number of instances for problem P .

Scoring

Each participant is awarded the sum of scores per each benchmark domain, naturally weighing NP (>50%) more than P (20-25%) and Beyond-NP/Opt (20-25%).

Instance Quota

$$S_{solve}(P) = \frac{N_S}{N}$$

N_S = number of solved instances for problem P , N = total number of instances for problem P .

Time Quota

$$S_{time}(P) = \frac{1}{N} \sum_{i=1}^N \left(1 - \left(\frac{\log(t_i + 1)}{\log(t_{out} + 1)} \right) \right)$$

t_{out} = maximum allowed time, t_i on instance i , N as above.

More on the time quota

Awards faster systems logarithmically. For $t_{out} = 600$ and S_{time} normalized to 50:

Time (sec)	Score (0-50)
0	50.00
0.9	45.00
2.6	40.00
5.8	35.00
11.9	30.00
23.5	25.00
45.4	20.00
87.1	15.00
166.1	10.00
315.9	5.00
527.8	1.00
562.7	0.50

Similar as the log-based score proposed for SAT Competition 2009: naturally tends to award significant time score only to very fast systems.

Optimization Quota

 $S_{opt} =$

- Inconsistent instances: a flat score of $\frac{1}{2N}$ for detecting it;
- Satisfiable instances:
 - A flat reward of $\frac{1}{4N}$ for finding a witness;
 - A flat reward of $\frac{1}{4N}$ for finding an optimal witness;
 - A quality reward of

$$\frac{1}{2N} e^{\frac{B_i - Q_i}{Q_i}}$$

depending on the distance of the solution cost Q_i from the best solution cost B_i for instance i .

More on Optimization Quota

Some additional score awarded to solutions within 3% of the *measured* optimal solution. Finding an optimum is strongly rewarded. For $N = 1$ and S_{opt} normalized within 0 and 50:

Quality Gap (%)	Score (0-50)
0	50
1	22
2	16
3	14
>4	13

Outline

- ① The Third ASP Competition
- ② System Track
- ③ Model and Solve Track
- ④ Comparisons

Driving Principles of the System Track

Driving Principles of the System Track

Language standardization

- Try to enlarge the number of acknowledged common constructs of ASP systems;

Driving Principles of the System Track

Language standardization

- Try to enlarge the number of acknowledged common constructs of ASP systems;

Compare systems in fixed conditions

- Fixed input problem encoding;
- Fixed default heuristics and internal settings;

From driving principles to rules

Rules

- Language fixed to ASP-Core (larger ASP-RfC draft as a proposal);
- Organizers provide ASP-Core encodings for each benchmark;
- Forbidden to look for syntactic aspects of problem encodings in order to trigger ad-hoc heuristics (e.g. predicate names).

From driving principles to rules

Rules

- Language fixed to ASP-Core (larger ASP-RfC draft as a proposal);
- Organizers provide ASP-Core encodings for each benchmark;
- Forbidden to look for syntactic aspects of problem encodings in order to trigger ad-hoc heuristics (e.g. predicate names).

Scoring

For benchmark P , a participant is awarded $S(P)$ according to:

$$S(P) = \alpha S_{solve}(P) + (100 - \alpha) S_{time}$$

Where $\alpha = 50$. Note that the logarithmic behavior of S_{time} naturally weighs this term less than $S_{solve}(P)$.

Participants

- **clasp**, **claspfolio** and **clasp**, Potassco team at University of Potsdam
- **CModels** and **SUP**, University of Kentucky
- **IDP**, KRR Group at KU-Leuven
- **lp2*** systems and **smodels**, Aalto University, former Helsinki UT

Polynomial Problems

Polynomial Problems

System	P	P-inst	P-time
clasp	213	150	63
claspfolio	209	150	59
claspd	206	145	61
sup	195	140	55
lp2sat2gminisat	185	140	45
cmodels	184	130	54
idp	184	130	54
smodels	180	130	50
lp2sat2minisat	179	135	44
lp2diffz3	178	135	43
lp2sat2lminisat	171	130	41

NP Problems

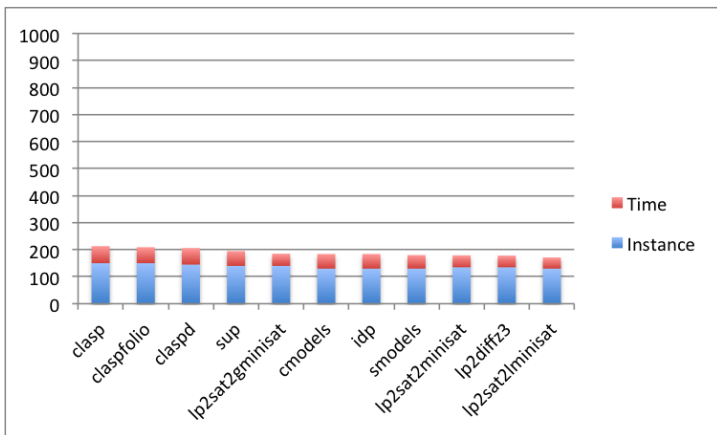
NP Problems

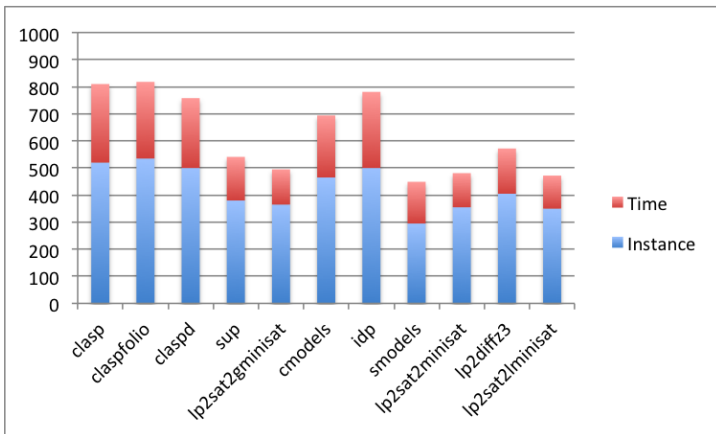
System	NP	NP-inst	NP-time
claspfolio	609	385	224
clasp	597	370	227
idp	597	370	227
claspd	552	355	197
cmodels	510	335	175
lp2diffz3	394	270	124
sup	346	240	106
lp2sat2gminisat	310	225	85
lp2sat2minisat	302	220	82
lp2sat2lminisat	301	220	81
smodels	269	165	104

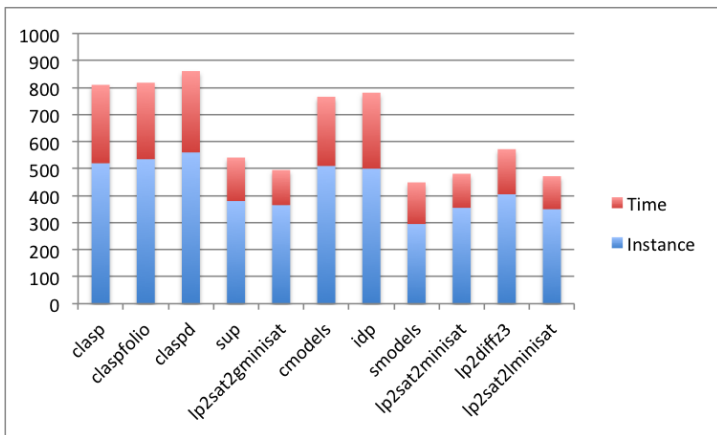
Beyond NP Problems

Beyond NP Problems

System	Beyond NP	Beyond NP-inst	Beyond NP-time
clasp	103	60	43
cmodels	72	45	27







Overall Scoring

Overall Scoring

System	Total	Inst	Time
claspd	861	560	301
claspfolio	818	535	283
clasp	810	520	290
idp	781	500	281
cmodels	766	510	256
lp2diffz3	572	405	167
sup	541	380	161
lp2sat2gminisat	495	365	130
lp2sat2minisat	481	355	126
lp2sat2lminisat	472	350	122
smodels	449	295	154

Outline

- ① The Third ASP Competition
- ② System Track
- ③ Model and Solve Track
- ④ Comparisons

Driving Principles of the Model and Solve track

To foster open comparison with any other declarative paradigm
Competition open to any “declarative”-based system

Driving Principles of the Model and Solve track

To foster open comparison with any other declarative paradigm

Competition open to any “declarative”-based system

To foster development of new linguistic constructs

No restriction on language. Participants develop their own problem specifications

Driving Principles of the Model and Solve track

To foster open comparison with any other declarative paradigm

Competition open to any “declarative”-based system

To foster development of new linguistic constructs

No restriction on language. Participants develop their own problem specifications

To foster development of new heuristics and/or algorithms

No restrictions in fine-tuning, on a per benchmark basis.

From driving principles to rules

Rules

Participants submit a “solution” on a per benchmark basis.
Knowledge of the benchmark domain can be exploited (yet not the knowledge of the instance family)

From driving principles to rules

Rules

Participants submit a “solution” on a per benchmark basis.
Knowledge of the benchmark domain can be exploited (yet not the knowledge of the instance family)

Scoring

- **Search and Query problems:** same as in the System track;
- **Optimization problems:** for problem P , a participant team is awarded $S(P)$, where

$$S(P) = \alpha S_{opt}(P) + (100 - \alpha) S_{time}$$

for $\alpha = 50$.

Participants

- **aclasp**, Potsdam University
- **BPSolver**, CU - New York, University of Udine, Texas Tech University
- **Ezcsp**, Eastman Kodak Company, University of Kentucky
- **fastdownward**, Universitaet Freiburg, Universidad Carlos III de Madrid, Technion Univ. NICTA, University of British Columbia
- **IDP**, KRR Group at KU-Leuven
- **Potassco** team at University of Potsdam

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
idp	175	117	58	148

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
fastdownward	181	100	81	181
idp	175	117	58	148

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
ezcsp	320	173	147	285
fastdownward	181	100	81	181
idp	175	117	58	148

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
aclasp	404	240	164	365
ezcsp	320	173	147	285
fastdownward	181	100	81	181
idp	175	117	58	148

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
bpsolver	459	253	206	459
aclasp	404	240	164	365
ezcsp	320	173	147	285
fastdownward	181	100	81	181
idp	175	117	58	148

Polynomial Problems

System	P	P-inst	P-time	(Pure) - P
clasp	497	290	207	459
bpsolver	459	253	206	459
aclasp	404	240	164	365
ezcsp	320	173	147	285
fastdownward	181	100	81	181
idp	175	117	58	148

NP Problems

System	NP	NP-inst	NP-time

NP Problems

System	NP	NP-inst	NP-time
fastdownward	105	70	35

NP Problems

System	NP	NP-inst	NP-time
bpsolver	1126	607	519
fastdownward	105	70	35

NP Problems

System	NP	NP-inst	NP-time
idp	1127	680	447
bpsolver	1126	607	519
fastdownward	105	70	35

NP Problems

System	NP	NP-inst	NP-time
aclasp	1224	680	544
idp	1127	680	447
bpsolver	1126	607	519
fastdownward	105	70	35

NP Problems

System	NP	NP-inst	NP-time
ezcsp	1419	786	633
aclasp	1224	680	544
idp	1127	680	447
bpsolver	1126	607	519
fastdownward	105	70	35

NP Problems

System	NP	NP-inst	NP-time
clasp	1481	848	633
ezcsp	1419	786	633
aclasp	1224	680	544
idp	1127	680	447
bpsolver	1126	607	519
fastdownward	105	70	35

Beyond NP

System	Beyond NP	Beyond NP-inst	Beyond NP-time

Beyond NP

System	Beyond NP	Beyond NP-inst	Beyond NP-time
bpsolver	86	43	43

Beyond NP

System	Beyond NP	Beyond NP-inst	Beyond NP-time
clasp	110	60	50
bpsolver	86	43	43

Optimization Problems

System	Opt	Opt-opt	Opt-time

Optimization Problems

System	Opt	Opt-opt	Opt-time
ezcsp	34	34	0

Optimization Problems

System	Opt	Opt-opt	Opt-time
fastdownward	81	47	34
ezcsp	34	34	0

Optimization Problems

System	Opt	Opt-opt	Opt-time
idp	140	121	19
fastdownward	81	47	34
ezcsp	34	34	0

Optimization Problems

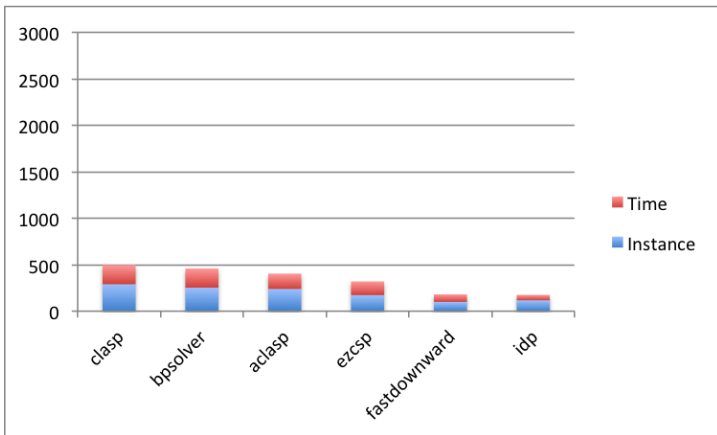
System	Opt	Opt-opt	Opt-time
bpsolver	207	161	46
idp	140	121	19
fastdownward	81	47	34
ezcsp	34	34	0

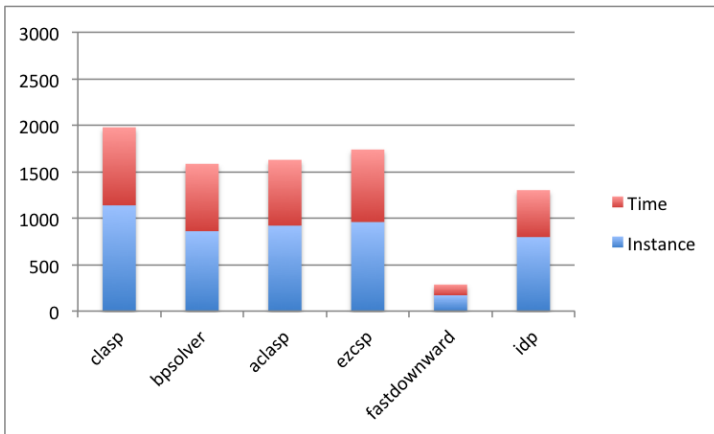
Optimization Problems

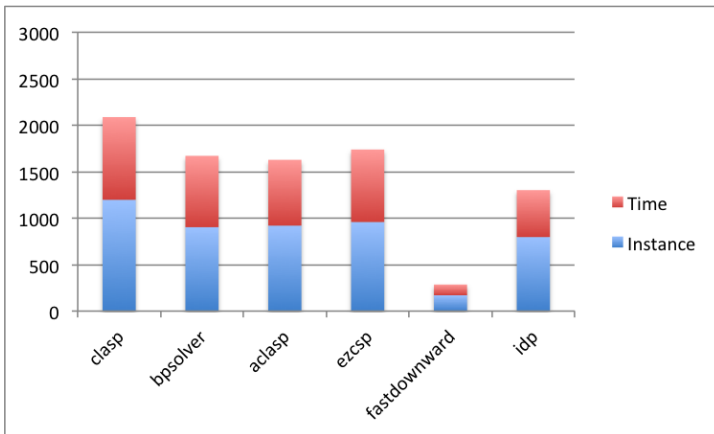
System	Opt	Opt-opt	Opt-time
aclasp	325	220	105
bpsolver	207	161	46
idp	140	121	19
fastdownward	81	47	34
ezcsp	34	34	0

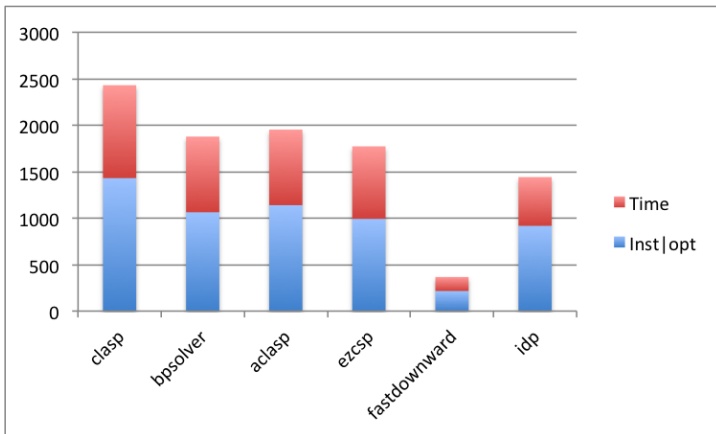
Optimization Problems

System	Opt	Opt-opt	Opt-time
clasp	343	234	109
aclasp	325	220	105
bpsolver	207	161	46
idp	140	121	19
fastdownward	81	47	34
ezcsp	34	34	0









Overall Ranking

System	Total	Inst	Time

Overall Ranking

System	Total	Inst	Time
fastdownward	367	217	150

Overall Ranking

System	Total	Inst	Time
idp	1442	918	524
fastdownward	367	217	150

Overall Ranking

System	Total	Inst	Time
ezcsp	1773	993	780
idp	1442	918	524
fastdownward	367	217	150

Overall Ranking

System	Total	Inst	Time
bpsolver	1878	1064	814
ezcsp	1773	993	780
idp	1442	918	524
fastdownward	367	217	150

Overall Ranking

System	Total	Inst	Time
aclasp	1953	1140	813
bpsolver	1878	1064	814
ezcsp	1773	993	780
idp	1442	918	524
fastdownward	367	217	150

Overall Ranking

System	Total	Inst	Time
clasp	2431	1432	999
aclasp	1953	1140	813
bpsolver	1878	1064	814
ezcsp	1773	993	780
idp	1442	918	524
fastdownward	367	217	150

Outline

- ① The Third ASP Competition
- ② System Track
- ③ Model and Solve Track
- ④ Comparisons

A few comparisons

REACHABILITY	Grammar-Based II		GRAPH COLORING		CROSSING MIN		FastFoodOpt		Maximal Clique		overall VS Sys		
											clasp	1198	
											ezcsp	1069	
bpsolver	90	bpsolver	98								bpsolver	1030	
clasp	87	aclasp	85	idp	66		clasp	83	clasp	64	aclasp	949	
idp	76	clasp	85	clasp	51		aclasp	83	aclasp	64	idp	885	
XSB 3.2	65	Clasp 2009	80	Clasp 2009	38	CPLEX 12	83	Clasp 2009	77	Clasp 2009	64	clasp (System)	701
						Minisat+	52			Cliquer 2.1	62		
				aclasp	28	clasp	17	idp	64	idp	58	fastdownward	302
				ezcsp	23	idp	13	bpsolver	48	bpsolver	50		
				bpsolver	15	aclasp	13						
						bpsolver	13						

Competition Figures

Competition Figures

- Estimated 25000 lines of new code
(C/C++/Python/Perl/Java/Prolog/ASP/Bash ...).

Competition Figures

- Estimated 25000 lines of new code (C/C++/Python/Perl/Java/Prolog/ASP/Bash ...).
- Benchmark Domains: 35 (18 for the System Track), a total of 506 instances and 2920 + 2123 runs, 1 week wall clock CPU time, excluding dry runs and testing

Competition Figures

- Estimated 25000 lines of new code (C/C++/Python/Perl/Java/Prolog/ASP/Bash ...).
- Benchmark Domains: 35 (18 for the System Track), a total of 506 instances and 2920 + 2123 runs, 1 week wall clock CPU time, excluding dry runs and testing
- Hardware: 4x4 core Intel Xeon CPU X3430 / 2.4 Ghz / 4GB RAM

Competition Figures

- Estimated 25000 lines of new code (C/C++/Python/Perl/Java/Prolog/ASP/Bash ...).
- Benchmark Domains: 35 (18 for the System Track), a total of 506 instances and 2920 + 2123 runs, 1 week wall clock CPU time, excluding dry runs and testing
- Hardware: 4x4 core Intel Xeon CPU X3430 / 2.4 Ghz / 4GB RAM
- Manpower: 3 Chairs, 16 among Ph.D. Students, Professors, Webmasters, Sysadmins

Competition figures

Competition figures

- Resource allocation: 1 Million Marble blocks; 2 Triremes; 4 Miles steel chains; More than 600 wooden oars; 20 wasted weekends;

