Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Automatic Configuration of Benchmark Sets for Classical Planning

Álvaro Torralba,¹ Jendrik Seipp,² Silvan Sievers² ¹Aalborg University, Denmark ²University of Basel, Switzerland

October 21, 2020

The	ICAPS	Way
000		

Benchmark Configuration

Evaluation

Conclusion

Outline



- 2 Benchmark Design Principles
- Benchmark Configuration

4 Evaluation



Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning (But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

Benchmark Configuration

Evaluation

Conclusion

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning (But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

Is, then, the life of a researcher in heuristic search planning characterized by the following pseudo-code?

Fig. 3. The life of a planning researcher?

Benchmark Configuration

Evaluation

Conclusion

The Cycle of Life (in Planning Research)

Everything you Always Wanted to Know About Planning (But Were Afraid to Ask) — (Jörg Hoffmann, 2011)

Is, then, the life of a researcher in heuristic search planning characterized by the following pseudo-code?

Fig. 3. The life of a planning researcher?

The answer to that one is "NO!". Far beyond just improving performance on benchmarks, the *understanding* of heuristics is where heuristic search planning really turns into a natural science. Dramatic progress has been made, in that science, during the last years. For example, Bonet and Geffner 3 proved

The	ICAPS	Way
000	0	

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co	verage	2		Time	Score	-
	_	2	.51			k	3	
	2	-8	- 13	8	8	- 2	- 13	8
	- 8	- 2	- 2	- 8	10	- 2	- 2	- 6
	0	0	0	0	0	0	0	
agricola (20)	0	٠	٠		0,0	0.0	0.0	0.0
airport (50)	- 34	24	- 34	34	2.5	18.5	26.1	26.1
barman (34)	- 4	0	- 4	- 4	0.3	0.0	2.0	1.8
blocks (33)	28	20	28	28	2.2	19.5	29,1	29.1
childsnack (20)	. 0	٠			0,0	0.0	0.0	0.0
data-network (20)	- 14		- 14	- 14	1.1	7.2	12.1	12.1
depot (22)	13	- 6	13	-13	1.0	- 3.3	9.6	9.8
driverlog (26)	15	- 7	15	15	1.1	- 4.3	10.5	10.7
elevators (50)	- 44	12	- 44	44	3.4	2.6	25.9	26.6
floortile (40)	- 6	0	•		0.3	0.0	0.8	0.8
freecell (88)	68	- 30	68	68	4.8	10.9	35.8	36.1
god (20)	19	- 7	19	19	1.5	- 4.7	12.1	12.2
grid (5)	3		- 3	- 3	0.2	1.0	2.2	2.3
Buibbes (30)		. 6			0.0	-4.8	7.0	6.8
hiking (20)	14	. 8	15	15	1.0	5.1	10.9	10.7
logistics (s)	39	19	39	39	2.8	11.8	25.2	26.0
miconic (150)	144	133	144	144	11.2	80.6	131.6	131.9
movie (50)	- 20		- 20	- 20	- 24	42.7	42.4	42.0
inprime (35)	29	-24	- 29	29	2.3	19.1	26.5	26.7
mystery (30)	- 19	12			1.0	12.6	17.8	17.8
nonrystery (20)	20	12	- 28	20	1.5	8.0	15.5	15.4
openstacks (100)	- 10	21	- 52	- 12	3.8	12.0	29.8	30.0
organic (20)		- 7	- 2		0.5	6.0		0.1
organic-split (20)	10			10	0.7	1.9		
burchannes (20)	- 28	- 24			2.9	28.2	34.1	34.2
parking (an)	- 13	- 1	-12	- 12	0.9	0.1	- M	- 54
pathways (30)		- 21			- 0.3	2.1		- 24
pegaol (30)	48	42	- 12		3.7	22.8	35.6	35.4
petri-net (20)								
biber-ur Got		12	- 22	45	- 12	10.4	19.0	10.7
pipes-1009	18				1.3	2.1	141	12.0
bit-mail (50)		- 22			- 22	10.4		54.5
FOTERS (43)		- 2	- 2	- 2	0.6	0.0		- 54
sateline (m)		- 2			0.3	4.7		
scanaryzer (50)	- 22	- 2	- 22			- 24	12.0	
snake (20)	- 10					10.6		- 10.0
souther care	- 14	1	- 77	- 22	- 11	2.0	122	- 22
spearer (20)	10				- 12	12.4		
somage (N)	- 13	12	- 65			100	111	111
testale calls	- 55		10	- 22	0.8	1.1		
tidadeot care	- 16	16	- 34		12	- 66	- 63	164
100 (10)	- 10	- 12	1	- 10	0.6	80	8.9	8.9
transport (30)	- 14	- 20	- n	- M	26	10.4	22.4	22.4
trucks con	- 13	19	- 13	- 13	0.9	53	8.6	8.8
visitall cars	30	33	30	30	2.3	27.8	30.3	30.3
woodwork on	49	28	49	49	3.8	24.3	40.5	44.7
zenotravel (20)	13	7	13	13	1.0	4.9	8.7	8.8
Same (1977)	1156	766	1155	1159	86.8	\$10 K	900.4	908.7
Comment (cond ()		1.00						10000

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co	verap			Time	Score	
		3				ξ.	ž	
		2		<u>, 8</u>		20	-	- 2
	- 9	- 8	- 8	- 8	- 8	- 8	- 8	- 18
seriesla cos	0				0.0	0.0	0.0	0.0
aimort (ND	- 34	24	- 34	- 34	2.5	18.5	26.1	26.1
barman (Mr.	14	10	14		0.3	0.0	2.0	1.8
blocks (15)	28	20	28	28	2.2	19.5	29.1	29.1
childenack (20)					0.0	0.0	0.0	0.0
data-network can	- 14	лī.	- 14	- 64	11	7.2	12.1	12.1
depot (22)	13	6	13	- 13	1.0	3.3	9.6	9.8
driverlog can	15	7	15	15	1.1	4.3	10.5	10.7
elevators con	- 44	12	44	44	3.4	2.6	25.9	26.6
floortile ons	6	0	- 6	- 6	0.3	0.0	0.8	0.8
freecell (80)	- 68	30	68	68	4.8	10.9	35.8	36.1
god (20)	19	7	19	19	1.5	4.7	12.1	12.2
grid (5)	- 3		- 3	3	0.2	1.0	2.2	2.3
aripper (20)	8	- 6		8	0.6	4.8	7.0	6.8
hiking (20)	14	8	15	15	1.0	5.1	10.9	10.7
logistics (67)	39	19	39	39	2.8	11.8	25.2	26.0
miconic (15th	144	133	144	144	11.2	80.6	131.6	131.9
movie can	- 30	30	- 30	30	2.4	42.7	42.4	42.6
mprime (35)	29	24	29	29	2.3	19.1	26.5	26.7
mystery (30)	19	15	19	19	1.5	12.6	17.8	17.8
nonrystery (20)	20	12	20	20	1.5	8.0	15.3	15.4
openstacks (100)	- 53	24	53	53	3.8	12.0	29.8	30.0
organic (20)	7	7	7	7	0.5	6.0	6.1	6.1
organic-split (26)	10	- 6	10	10	0.7	1.9	4.2	4.2
parcprinter (50)	- 38	34	38	38	2.9	28.2	34.1	34.2
parking (40)	13		13	13	0.9	0.1	5.4	5.4
pathways (50)	. 8	- 4	- 8		0.3	5.1	8.8	- 5.4
pegsol (80)	-48	42	-48	48	3.7	22.8	35.6	35.4
petri-net (20)	. 0	٠			0,0	0.0	0,0	0,0
pipes-m (50)	25	14	25	25	1.8	10.4	19.0	18.7
pipes-1 (50)	18	8	18	18	1.3	5.1	12.1	12.0
per-small (20)	50	-49	50	50	3.9	48.2	54.6	54.5
EDTEES (43)	8	7			0.6	6.6	8.1	8.1
satellite (36)	7	6	7	7	0.5	5.5	7.3	7.2
scanalyzer (M)	35	7	33	35	2.7	5.7	19.6	21.2
snake (20)	12	.6	12	12	0.9	2.5	7.6	7.4
sokoban (50)	50	33	.50	50	3.8	19.6	39.5	39.9
spider (20)	15	2	15	15	1.1	2.9	8.6	8.5
storage (20)	16	14	16	16	1.2	12.5	17.1	17.1
lermes (20)	12	0	12	12	0.8	0.0	3.2	3.2
tetris (17)		3	10		0.8	1.3	5.4	5.5
tidybot (#I)	25	18	- 24	25	1.8	5.8	15.3	15,4
tpp cm	8	7			0.6	8.0	8.9	8.9
transport (20)	- 24	20	- 36	36	2.6	10,4	22.4	22,4
trucks (20)	13	.2	- 63	- 63	0.9	5.3	8.6	8.8
visital (40)	30	33	30	30	2.3	27.8	30.3	30.3
woodwork (50)	- 49	38	- 42	- 42	3.8	24.3	40.5	44.7
senotravel (20)	13	7	-13	13	1.0	4.9	8.7	8.8

					L (Mp)					
	N ⁱⁿ	a	я	-	DNS	g	0	3	τ	
agricola	0	0	10	16	1	13	12		9	
airport.	23	- 35	- 36	- 33	1	21	- 13	- 14	17	
bornan	0	0	12	20	0	0	0	0	0	
blocks	18	35	35	35	21	33	27	35	35	
childenack	0	0	- 1	0	0	0	0	0	0	
data-network	0	- 1	5	2	0	0	0	0	0	
depot	- 4	14	18	18	6	- 5	- 5	1.5	13	
driverlog	2	19	18	18	8	12	- 13	15	- 14	
elevators		- 5	20	. 7	0	0	0	0	0	
footile	0	0	2	1	0	0	0	0	0	
freecell	20	-46	79	- 88	18	50	23	- 61	57	
pod	0	20	20	20	0	20	0	0	0	
prid	1	- 3	- 4		0	- 3	- 0	- 3	- 4	
proper	. 8	20	20	20	8	20	10	20	20	
biking	- 2	- 3	20	20	2	- 7	- 5	- 3	- 3	
logistics	- 2	7	29	15	2	2	- 3	6	- 5	
maintenance	•	14	11	14	0	0	0	0	0	
miconic	55	150	170	150	71	150	145	150	150	
movie	30	- 36	- 34	- 34	30	- 30	- 30	- 39	- 30	
reprint	20	21	33	22	5	18	13	18	17	
reptoy	15	15	17	14	9	- 6	12	10	9	
openstacks	0	0	- 2	20	0	1	8	2	5	
organic-symbolis	- 3	- 3	- 2	- 3	1.8	3	- 3	- 3	2	
parcprinter		12	20	18	0	0	1	0	0	
parking		0	. 7	0	0	0	0	0	0	
pathways	- 4	5	10		4	- 4	- 4	- 4	- 4	
pegsel	17	20	20	20	18	20	20	20	20	
piperworld	12	22	23	27	12	13	15	20	17	
per	49	50	54	50	50	50	50	50	50	
revers	6	21	26	- 25	6	- 14	- 16	16	16	
satellity	6	15	27	12	7	15	. 8	. 8	9	
scanalyzer	- 4	20	18	20	5	18	11	20	18	
seake	3	- 4	5	7	4	3	3	- 7	6	
soloban	6	13	19	10	<u>к</u>	11	12	10	9	
spider	1	12		19	0	- 5	- 7	- 5	- 9	
storage	34	18	12	19	17	15	19	18	19	
NOTION .	0	10	14	15	3	- 4	2	- 7	- 3	
tetris.	0	20		20	2	11	19	2	18	
freightfall	- 5	. 5		14	5	12	5	5	5	
tidabot	- 3	19	16	20	0	1	- 7	8	2	
100	- 6	13	- 23	- 29	6	15	10	14	15	
transport .	0	5	•	16	0	0	0	0	0	
trucks	- 6		15	. 9	6	6	7	- 7	7	
visital		20	•	20	0	0	. 9	0	0	
woodworking	1	1	- 2	4	1	1	1	1	1	
constravel	8	20	20	20	7	- 9	- 9	13	14	
SUM	359	775	933	968	346	651	558	631	632	
$>h_E$					4	6	4	6		
>hg					2	5	5	- 4	- 3	
>0-					0	2	1	0	0	

1

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co				Time	Score			
			tring.		_		accer	_		-
		- 6	- 2			- 4	ž			
	_ <u>3</u>	2	-0	.2	iic,	20	-	<u> 8</u>		
	- 9	- 8	- 8	18		- 8	- 8	18		
and a data see	-									
agriceta (20)										-
anperceso	- 22	- 12	- 22	- 22	- 53	18.0	-	-	agricola	
burnan (50)						0.0	101		airport	
DOCKS COS	- **	- 22	- 12		- 22	19.0	-	-	borman	
CHIRDNERCK (20)	- 44								blocks	
CALL DETIVORY (20)	- 23	- 12	- 53	- 53	- 11	- 22	100	10.0	childrack	
depot (22)	- 22		1.2	10	1.0	- 23	9.0		data-external	
diversity (20)	- 22		- 55		- 22	- 22	10.5	10.1	4000	
clevators (50)	. **	12			- 22	2.0	25.9	20.0	delegation	
Boortine (40)		.0			0.3	0.0	0.8	0.8		
reccen (80)	- 65	- 20	- 68	- 68	4.8	10.9	25.8	20.1	COVIER D	
Borg (3a)	- 12	- 7	19	19	1.5	4.7	12.1	12.2	Postav	
grid (5)	- 2	. !	- 2		0.2	1.0	- 22	2.3	Deecell	
Buddhese (56)		0			0.6	4.8	7.0	0.8	ped	
hiking (20)	- 14		15	15	1.0	5.1	10.9	10.7	prid	
logistics (67)	39	19	39	39	2.8	11.8	25.2	26.0	Dipper	
meconec (150)	144	123	144	144	11.2	80.6	131.6	131.9	billing	
movile (30)	- 30	- 20	- 50	50	- 24	42.7	424	42.6	logistics	
imprime (35)	29	24	29	19	2.3	19.1	26.5	26.7	maintenance	
mystery cos	19	15	19	19	1.5	12.6	17.8	17.8	micronic	
nonystery (30)	20	12	20	20	1.5	8.0	15.3	15.4	movie	
openstacks (10)	- 53	21	- 53	53	3.8	12.0	29.8	30.0	and the second second	
organic (33)	7	7	7	7	0.5	6.0	6.1	6.1	motory	
organic-split (20)	10	- 6	10	10	0.7	1.9	4.2	4.2	operator by	
parcprinter (50)	38	-34	38	38	2.9	28.2	34.1	34.2	cyclosed as	
parking (at)	-13		13	13	0.9	0.1	- 5.4	5.4	organic-system	1
pathways (30)	- 8	- 4	8	5	0.3	5.1	5.5	5.4	Indente	
pegsel (50)	-48	42	48	48	3.7	22.8	35.6	35.4	barried	
petri-net (20)	٠	0	0	0	0.0	0.0	0.0	0.0	pathways	
pipes-mi (50)	25	-14	25	2.5	1.8	10.4	19.0	18.7	belies	
pipes-t (50)	18	8	18	18	1.3	5.1	12.1	12.0	piperworld	
par-small (50)	50	49	50	50	3.9	48.2	54.6	54.5	let.	
FOTEES (40)	8	- 7	8	8	0.6	6.6	8.1	8.1	revers	
satellite (36)	7	6	7	7	0.5	5.5	7.3	7.2	unilla	
scanalyzer (50)	35	7	33	35	2.7	5.7	19.6	21.2	sconalyzer	
snake (20)	12	6	12	12	0.9	2.5	7.6	7.4	seake	
sokoban (50)	50	33	50	50	3.8	19.6	39.5	39.9	solution	
spider (20)	15	7	15	15	1.1	2.9	8.6	8.5	spider	
storage (36)	16	14	16	16	1.2	12.5	17.1	17.1	sharage	
lermes (20)	12	0	12	12	0.8	0.0	3.2	3.2	NOTION .	
tetris (17)	- 11	- 3	10	11	0.8	1.3	5.4	5.5	bittin.	
tidybot (40)	25	18	24	25	1.8	5.8	15.3	15.4	Desideful	
tpp cm	. 8	7	8	8	0.6	8.0	8.9	8.9	idebra .	
transport (70)	- 34	20	36	36	2.6	10.4	22.4	22.4	they been	
trucks (70)	13	- 9	13	13	0.9	5.3	8.6	8.8	46	
visital1 (m)	- 30	33	30	30	2.3	27.8	30.3	30.3	transport	
woodwork (50)	49	38	49	49	3.8	24.3	40.5	44.7	Inacks	
zenotravel (20)	13	7	13	13	1.0	4.9	8.7	8.8	vicial	
Parent contract.		76.6		1170		# 70. W	000.4	008.7	woodworking	
oun (162)	1120	/00	1120		- 60.8	119.9	7181.4	2001	centravel	
									SUM	
									$> h_R$	

					LONG				
	Nad	8	5	-	SNOT	a	0	5	υ
agricola	0	0	10	16	1	13	12		9
airport	23	- 35	- 36	- 33	1	21	13	- 14	17
borman	0	0	12	20	0		0	0	- 0
blocks	15	35	35	35	21	33	27	35	35
childonack	0	0	- 1	0	0		0	0	0
data-network	0	1	5	2	0		0	0	0
depet	- 4	- 14	18	18	- 6	- 5	- 5	1.5	13
driverlog	- 7	- 19	18	18	- 8	12	13	15	- 14
devaters	0	- 5	- 20	- 7	- 0	•	0	0	- 0
footile	0	0	2	1	0		0	0	0
freedl	- 30	-45	79	80	15		23	- 61	57
god	0	20	20	20	0	29	0	0	0
prid	1	- 3	- 4	- 8	0	- 3	0	- 3	- 4
preper	- 8	20	20	20	- 8	38	00	29	- 20
biking	2	- 3	20	20	2	- 7	- 5	- 3	- 3
logistics	2	- 7	29	15	2	2	- 3	6	5
maintenance	0	14	11	14	0		0	0	0
miconic	35	150	150	150	71	170	145	150	150
movie	30	- 39	- 30	- 30	- 30	30	- 30	- 39	- 30
mprime	-30	21	32	22	- 5	18	13	18	17
maker	15	1.5	87	14	- 9	- 6	12	10	- 9
openstacks	0	0	2	20	0	1.1	8	2	
organic-symbolis	3	- 3	2	- 3	3	3	3	- 3	2
percerinter	0	12	- 20	18	0	•		0	0
rocking	0	0		0	0	•	0	0	0
pathways	4	- 5	- 80		4	4	- 4	- 4	- 4
pegad	17	28	20	20	18	28	28	29	20
piperworld	12	22	23	27	12	13	15	30	17
Def.	43	59	.50	50	50	54	50	50	.50
revers	- 6	21	- 26	-25	- 6	- 14	16	16	- 16
satellite	- 6	15	27	12	- 7	15			- 9
scanalyzer	- 4	29	18	20	- 5	18	11	29	18
wake	3	- 4	5	7	- 4	3	- 3	- 7	- 6
solution	- 6	13	19	10	8	11	12	10	- 9
spider	1	12	- 9	19	0	- 5	- 7	- 5	- 9
storage	- 14	18	19	19	17	15	12	18	19
tormes.	0	10	14	15	- 3	- 4	2	- 7	- 3
tenis.	0	20	9	20	2	11	19	2	18
thoughtful	5	5	8	14	5	12	5	5	5
Edabot	- 3	19	- 16	20	0	1	- 7	- 8	2
100	- 6	13	-23	29	- 6	15	90	14	15
Biansport .	0	5	0	16	0		0	0	0
tracks	- 6	- 9	15	9	6	- 6	2	- 7	7
visital	0	29	0	20	0	0	. 9	0	0
woodworking	1	1.1	2	4	1	1	1	1.1	
constrained	8	20	20	20	2		- 9	13	14
SUM	399	775	933	965	345	651	558	631	632
$>h_{K}$					- 4	- 6	4	6	
>bg					2	- 5	- 5	- 4	- 3
$>b_{in}$					0	2	1	0	0

Coverage	arb	inv	rnd	GZD	BD	ZCA	VDM	ZCP	AM	GZD+BD	Scorpion
Airport (50)	28	27	23	29	27	28	24	24	27	28	29
Blocks (35)	28	27	28	28	28	28	28	28	28	28	28
DataNetwork (20)	12	12	12	13	12	12	12	12	12	13	14
Depot (22)	7	7	7	7	7	7	7	7	7	10	13
DriverLog (20)	13	14	13	14	13	13	13	13	13	13	15
Elevators (30)	22	22	20	22	22	22	22	22	22	22	24
Freecell (80)	15	15	15	24	16	12	15	15	21	33	64
Grid (5)	2	2	1	2	2	2	2	2	2	2	3
Hiking (20)	10	10	9	10	9	8	9	9	10	9	14
Logistics (63)	27	27	25	27	25	25	25	25	25	27	34
Mprime (35)	23	25	22	23	23	23	22	22	25	24	31
Mystery (19)	16	17	15	17	17	17	17	17	17	17	19
Nomystery (20)	16	16	14	17	15	14	18	18	16	18	20
Openstacks (80)	31	31	- 31	- 31	- 31	30	31	31	31	31	34
OrgSynth-split (20)	15	15	14	15	14	10	15	15	15	15	10
Parcprinter (30)	19	22	19	22	19	19	22	22	18	20	30
Parking (40)	9	9	6	9	10	10	12	12	8	13	13
Pegsol (36)	35	34	33	35	34	34	35	34	34	35	35
Pipes-notank (50)	18	18	17	18	17	17	18	18	17	18	25
Pipes-tank (50)	12	12	10	12	11	9	12	12	12	12	18
PNetAlignment (20)	9	. 9	7	9	9	9	9	9	- 9	9	0
Rovers (40)	9	11	9	9	9	9	9	9	9	9	9
Satellite (36)	8	12	7	8	14	13	15	15	10	14	8
Scanalyzer (30)	16	16	16	16	15	14	16	16	16	16	18
Snake (20)	6	6	- 4	6	- 4	- 4	6	6	6	7	13
Sokoban (30)	30	29	30	- 30	30	30	30	30	30	30	30
Spider (20)	111	11	9	12	- 11	9	- 11		10	12	15
Termes (20)	7	6	6	7	6	6	7	7	6	7	13
Tidybot (40)	23	22	20	23	20	15	22	22	22	23	22
VisitAll (40)	16	15	17	15	- 36	36	36	36	14	36	30
Woodworking (30)	19	22	19	19	20	22	20	20	22	20	30
Zenotravel (20)	13	13	12	13	12	12	13	13	13	12	13
Others (601)	331	331	331	331	331	331	331	331	331	331	346
Sum (1672)	856	865	821	873	869	850	884	883	858	914	1020

4/25

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co	verage			Time	Score	
		2				×.		
	8	- 2	- 12	2	8	- 2	- 18	2
	- 6	- 2	- 1	- 4	- 6	- 2	- 2	- 6
	- 8	8	8	8	8	8	8	8
sericola cos		0	0	0		0.0	0.0	0.0
aimort con	- 34	24	- M	- M	2.5	18.5	26.1	26.1
barman (ba	- 14	0	- 4		0.3	0.0	2.0	18
Marke out	- 16	20	- 16	- 16	2.2	19.5	19.1	29.1
childsnock cm						0.0	0.0	0.0
data-network chin	- 64	- ñ.	- 14	- 14		1.2	12.1	12.1
depend only	- 63	12	- 63	- 63	10	- 6.5		
the part (and)	- 52		- 12	- 12	- 17	- 22	10.5	10.7
divering the	- 22	- 16	- 22		- 53	- 32	16.0	36.6
dependent (100	- 72	12	- 72	- 72	- 63		100	
DOCTOR (M)						10.0		
treecen (80)		- 20			- 22	10.9	20.6	38.1
Berr (10)		- 1	- 12	- 12	10	- 25	10.1	100
Burg (2)	- 2	. !			0.2	- 10	- 22	2.3
East-down costs	- 2				0.6	- 22	7.0	0.8
reading (26)	- 11		15	15	1.0	3.1	10.9	10.7
togrames (63)		19			2.8	11.8	25.2	26.0
miconic (150)	144	130	144	144	11.2	80.0	131.6	131.9
movie (30)		20	- 20		- 24	42.7	42.4	42.0
reprinte (35)	- 29	24	29	29	2.3	19.1	26.5	26.7
mystery cos		15	19	19	1.5	12.0	17.8	17.8
nonystery (30)	- 20	12	20	20	1.5	8.0	15.3	15.4
openstacks (100)	- 83	21	10	10	3.8	12.0	29.8	30.0
organic (30)	- 7	- 7	- 7		0.5	6.0	6.1	6.1
organic-split (20)	10	6	10	10	0.7	1.9	4.2	4.2
pareprinter (50)	38	34	38	38	2.9	28.2	34.1	34.2
parking (at)	-13		13	13	0.9	0.1	5.4	5.4
pathways (50)	- 8	- 4	5	8	0.3	5.1	5.5	5.4
pegael (50)	-48	42	48	48	3.7	22.8	35.6	35.4
petri-net (20)	٠	0	0	0	0.0	0,0	0.0	0.0
pipes-at (50)	25	14	25	25	1.8	10.4	19.0	18.7
pipes-1 (5t)	18	- 8	18	18	1.3	5.1	12.1	12.0
per-small (50)	50	49	50	50	3.9	48.2	54.6	54.5
FOTOTS (40)		7	8	8	0.6	6.6	8.1	8.1
satellite (36)	. 7	6	7	7	0.5	5.5	7.3	7.2
scanalyzer (80)	35	7	33	35	2.7	5.7	19.6	21.2
snake (20)	12	6	12	12	0.9	2.5	7.6	7.4
sokohan (50)	50	33	50	50	3.8	19.6	10.5	39.9
spider (20)	15	7	15	15	1.1	2.9	8.6	8.5
MOTION COL	16	14	16	16	1.2	12.5	17.1	17.1
lermes (20)	12	0	12	12	0.8	0.0	3.2	3.2
Metris (17)	11	3	10	ii.	0.8	1.3	5.4	5.5
tids bot cam	25	18	24	25	1.8	5.8	153	15.4
top (http	1	12	12		0.6	8.0	8.9	8.9
transport (20)	- 34	20	36	36	2.6	10.4	22.4	22.4
trucks cars	- 13	- 9	- 13	13	0.9	5.3	86	8.8
visital um	30	×.	10	10	2.3	22.8	30.3	30.3
woodwork con	- ñ	38	- 69	49	3.8	24.3	40.5	44.7
nenotrangl chin	- ñ	1	- 13	- íú	1.0	4.9	87	8.8
and the second s					1.00	- 10	3.7	
Sum (1827)	1156	766	1155	1159	\$6.8	539.8	900.4	908.7
	_		_		_	_		

					(0Mp)				
	Nind	4	я	-	SNGL	a	0	5	ΰ
agricola	0	0	10	16	1	13	12		9
airport	23	35	- 36	- 33	1	21	13	14	17
borman	0	0	12	20	0		0	0	- 0
blocks	15	35	35	35	21	33	27	35	35
childmack	0	0	- 1	0	0		0	0	0
data-network	0	1	5	2	0		0	0	0
depet	- 4	- 14	18	18	- 6	- 5	- 5	1.5	13
driverlog	- 7	- 19	18	18	- 8	12	13	15	- 14
devators	0	- 5	- 20	- 7	- 0	•	0	0	- 0
footile	0	0	2	- 1	0	•	0	0	0
freecell	- 20	-45	- 79	80	15		23	- 61	37
ged	0	20	20	20	0	29	0	0	0
grid	1	- 3	- 4	- 8	- 0	- 3	0	- 3	
Existen	8	- 24	- 20	20	- 8	20	10	29	20
hiking	- 2	- 3	20	20	2	2	- 5	- 3	
logistics	- 2	- 2	29	15	- 2	- 2	- 2	- 6	- 5
maintenance	0	- 14		- 14	0	•	0	0	0
miconic	-55	150	150	150	71	150	145	150	150
monie	39	- 34	- 39	- 30	39	34	- 34	- 34	- 30
ngeine	- 30	21	32	- 22	- 5	18	13	18	17
mystery	15	15	87	- 14	- 9	- 6	12	10	. 9
openstacks	0	0	2	20	0	1		- 2	- 5
organic-synthesis	3		- 2	- 3	3	3			- 2
porcprinter	0	12	- 20	18	- 0	•		- 0	0
parking	0	0	- 7	0	- 0	•	0	- 0	0
pothways	- 4	- 5	60		- 4	- 4	- 4	- 4	
pegad	17	29	20	20	15	20	29	29	20
piperworld	12	22	23	27	12	13	15	20	17
per .	49	59	50	50	50	20	59	59	50
revers		- 21	- 26	- 25	2	14	15	15	16
sate Bao		15	- 27	12	2	15			
scanalyzer	- 4	29	18	20	- 5	18		29	18
veake	3	- 4	- 5	- 7	- 4	3	3	- 7	6
sobobun	6	13	19	10	8	11	12	00	
spider	1	12	- 2	19	0	- 5	- 2	- 5	- 2
storage	14	18	19	19	17	15		18	19
termes.	0	10	- 14	15	1.2	1	2	- ?	- 2
MES	0	- 29	- 2	20	2		- 19	- 2	18
thoughtful		- 5		- 14	- 5	12	- 5	- 5	- 5
Edybox.	1.3	12	10	- 20			- 2		
100	. 6	- 13	- 23	29		15	99	- 14	15
transport	0	- 5	0	16	0	•	0		0
Inacks	6	. 2	15	. 2	6		- 2	- 2	- 7
visital	0	20	0	20	0		2	0	
woodworking		1	2	- 4				1	1
centravel	8	29	20	20	1.2			.19	
SUM	399	775	933	965	345	651	358	631	632
> h _g					1.1	1.1	- 1	. 1	- 7
>04					2	1.5	- 5	- 1	- 3



Coverage	arb	inv	rnd	GZD	BD	ZCA	VDM	ZCP	AM	GZD+BD	Scorpion
Airport (50)	28	27	23	29	27	28	24	24	27	28	29
Blocks (35)	28	27	28	28	28	28	28	28	28	28	28
DataNetwork (20)	12	12	12	13	12	12	12	12	12	13	14
Depot (22)	7	7	7	7	7	7	7	7	7	10	13
DriverLog (20)	13	14	13	14	13	13	13	13	13	13	15
Elevators (30)	22	22	20	22	22	22	22	22	22	22	24
Freecell (80)	15	15	15	24	16	12	15	15	21	33	64
Grid (5)	2	2	1	2	2	2	2	2	2	2	3
Hiking (20)	10	10	9	10	9	8	9	9	10	9	14
Logistics (63)	27	27	25	27	25	25	25	25	25	27	34
Murime (35)	23	25	22	23	23	23	22	22	25	24	31
Mystery (19)	16	17	15	17	17	17	17	17	17	17	19
Nottwstery (20)	16	16	14	17	15	14	18	18	16	18	20
Operatacks (80)	31	31	31	31	31	30	31	31	31	31	34
OrgSynth-split (20)	15	15	14	15	14	10	15	15	15	15	10
Parcerinter (30)	19	22	19	22	19	19	22	22	18	20	30
Parking (40)	9	9	6	9	10	10	12	12	8	13	13
Presol (36)	35	34	33	35	34	34	35	34	34	35	35
Pines-notank (50)	18	18	17	18	17	17	18	18	17	18	25
Pipes-tank (50)	12	12	10	12	- 11	0	12	12	12	12	18
PNetAlignment (20)	9	9	2	9		9	9	9		9	0
Rovers (40)	9	- 11	9	9	9	9	9	9	9	9	9
Satellite (36)	8	12	7	8	14	13	15	15	10	14	8
Scanalyzer (30)	16	16	16	16	15	14	16	16	16	16	18
Snake (20)	6	6	4	6	4	- 4	6	6	6	2	13
Sokoban (30)	30	29	30	30	30	30	30	30	30	30	30
Spider (20)	l ii	- îi	9	12	11	9	11	11	10	12	15
Termes (20)	7	6	6	7	6	6	7	7	6	7	13
Tidybot (40)	23	22	20	23	20	15	22	22	22	23	22
VisitAll (40)	16	15	17	15	36	36	36	36	14	36	30
Woodworking (30)	19	22	19	19	20	22	20	20	22	20	30
Zenotravel (20)	13	13	12	13	12	12	13	13	13	12	13
Others (601)	331	331	331	331	331	331	331	331	331	331	346
Sum (1672)	856	865	821	873	869	850	884	883	858	914	1020

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

Picocrilie 40 Freecell 42 GED 20 Grid 5 Logistics 63 Miconic 145 NoMystery 20 OpenSt14 20 PSR 48 Rovers 40 Satellite 36 Tidybot14 10 Transport 28 Woodwork 26 Zeastravel 20

		Co	verag			Time	Score	
		2	-8			2	-8	
	8	- 8	- 13	8	8	- 8	- 13	8
	- 61	- 2	- 1	- 2	- 6	- 7	- 2	- 2
	8	8	8	8	8	8	8	8
agricola (20)		0	0	0	0.0	0.0	0.0	0.0
airport con	- 34	24	- 34	34	2.5	18.5	26.1	26.1
barman (3-0	- 4	0	- 4	- 4	0.3	0.0	2.0	1.8
blocks (35)	28	20	28	28	2.2	19.5	29.1	29.1
childsnack (20)		0	0	0	0.0	0.0	0.0	0.0
data-network (20)	- 14	11	14	14	1.1	7.2	12.1	12.1
depot (22)	13	6	13	13	1.0	3.3	9.6	9.8
driverlog cm	15	7	15	15	1.1	4.3	10.5	10.7
elevators (70)	- 44	12	- 44	44	3.4	2.6	25.9	26.6
floortile cars	- 6	0	6	6	0.3	0.0	0.8	0.8
freecell on	68	30	68	68	4.8	10.9	35.8	36.1
god (20)	19	7	19	19	1.5	4.7	12.1	12.2
prid (5)	- 3	1	3	3	0.2	1.0	2.2	2.3
entiment (20)		6	8	8	0.6	4.8	7.0	6.8
hiking (20)	- 14	8	15	15	1.0	5.1	10.9	10.7
logistics (67)	39	19	39	39	2.8	11.8	25.2	26.0
miconic (150)	144	133	144	144	11.2	80.6	131.6	131.9
movie cm	30	30	30	30	2.4	42.7	42.4	42.6
movine (35)	29	24	29	29	2.3	19.1	26.5	26.7
members (20)	19	15	19	19	1.5	12.6	17.8	17.8
DOTTOMETY (20)	20	12	20	20	1.5	8.0	15.3	15.4
openstacks cross	53	21	53	53	3.8	12.0	29.8	30.0
organic con	2	7	7	7	0.5	6.0	6.1	6.1
organic-split on	10	6	10	10	0.7	1.9	4.2	4.2
percerinter (http://	38	- 34	38	38	2.9	28.2	34.1	34.2
metaling cats	13	1	13	13	0.9	0.1	5.4	5.4
outhways on	- 8	- 4			0.3	5.1	5.5	5.4
mental city	- 45	42	48	48	3.7	22.8	35.6	35.4
petri-pet con-					0.0	0.0	0.0	0.0
Dines at the	25	14	25	25	1.8	10.4	19.0	18.7
pipes-1 (56)	10	12	18	18	- 13	41	12.1	12.0
personal care		- m		50	1.0	48.2	14.6	54.5
per canada (Act	1	1	100	1	0.6	66	11	81
potelline cher	- 2	- 6			0.5	4.4	- 23	2.2
same the cost					3.7	4.7	10.6	
scaraby der (50)	- 11	- 6	- 66	- ñ	- 6.9	- 24	2.6	14
and advantage of the				-		10.6	10.5	10.0
sourceast (Sta	- 77	12	- 16		- 11	2.0	122	4.5
spager (20)	- 22	- 64	10	10	- 13	12.4		17.1
somage (30)	- 13	12	10	- 63	0.8	12.5	122	100
And all a second second	- 11		10	11	0.0		- 22	
tide box care		16	24	- 14	1.8	- 60	161	164
they can (80)		- 12	- 12	- 1			100	
transport chin	- ů	20			2.6	10.4	22.4	22.4
tracks can	- 6	- 0	- 65	- 61	0.9	5.3	2.6	
stated at 1,000	- 52		10	10	3.3		10.1	10.1
montheast (10)	- 22	10	49	10	1.0	24.2	40.5	44.7
WOODWORK (50)	- 22	- 12	- 77		- 10	24.5		
2000 and 1 (20)	- 10		13	- 10	1.0	4.9	6.7	
Sum (1827)	1156	766	1155	1159	\$6.8	539.8	900.4	908.7
		_		_	_		_	_

	Mind	14	ht	ulu	SNGL (Mp)	a	0	54	10
agricola	0	0	10	16	1	13	12		5
airport	23	- 35	- 36	- 33	1	21	13	- 14	17
borman	0	0	12	20	0		0	0	- 0
blocks	15	35	35	35	21	33	27	35	35
childmack	0	0	- 1	0	0		0	0	0
data-network	0	- 1	5	2	0		0	0	0
depet	- 4	14	18	18	- 6	- 5	- 5	1.5	13
driverlog	2	19	18	18	8	12	13	15	14
elevators	0	- 5	- 20	- 7	0		0	0	- 0
footile	0	0	2	1	0		0	0	0
freecell	- 30	-45	- 79	80	15	88	23	61	57
pod	0	20	20	20	0	28	0	0	0
grid	1	- 3	- 4	- 8	0	- 3.	0	- 3	- 4
ariterer .	8	28	- 20	20	8	24	- 10	20	- 20
biking	2	- 3	20	20	2	2	5	- 3	- 3
logistics	2	- 7	29	15	2	2	- 3	- 6	5
maintenance	0	14	11	14	0		0	0	0
miconic	35	159	150	150	21	159	145	159	150
movie	30	34	30	30	30	34	38	38	30
received	30	21	32	22	5	18	- 13	18	17
maker	15	15	17	14	- 9	6	12	00	
openstacks	0	0	2	20	0	1.1	8	2	
pressie combrais			- 2						2
percentator	0	12	- 20	18	0		- î	- 0	- 0
parking	0	0		0	0		- 0	0	0
reduces	1		- 10	- i	1	1	- 1	- 1	
period	12	- 28	20	- 20	18	28	- 28	20	20
nicecould	12	22	23	27	12	11	15	20	17
DI	43	50	50	- 50	50		53	50	
revera .	6	21	26	25	6	14	16	16	16
undate	6	15	27	12	2	15			9
scondurar.		20	1.0	20		18	- 11	- 20	10
under	1	1		- 7	1	1		- 7	6
schoburg	6	- 13	19	10		11	12	10	
unider	i.	12	- 9	19	0	5	2	5	- 6
storage	14	18	19	19	12	15		18	19
ALC: NO.	0	10	14	15		1	- 0	2	- 1
hada.		20		20	5	11	- 12		
foundated		- 1	- 6	14	- A	12	1	- 6	
tidebre .	3	19	16	20	0	1	- 2		2
box 1	6	13	23	29	6	15	- 22	- 14	15
Name of Street o				16					
tracks	6	- 9	15		6	6	- 2	- 2	7
visited	0	- 20		- 20					
materia	L ĭ	1	2	1	L ĭ	1.1	- 1	ĩ	ĩ
armstrand	8	- 28	- 20	- 20	2	14	- 6	- 6	- 14
SI M	199	125	***	-	146	651	558	611	612
3he					4	6		6	
>ba					2	5	- 5	- 4	
24					1.6	1.5	- 1		
(******									



			_		_	_		_													
	Des		B	tind 2	iearch	Deal		<u>.</u>	1.000		Duni	A									
	11	-117	377	-	-117	- ap	1					all off									্প
2	×	27.	$\supset d$	20	20	30	20	$\supseteq h$	$\supset D$	24	28 2	<u>a ⊃a</u>								<u>o</u> .	8
	1	2	- 1	- 2	- 1	1.3	- 2	- 2	- 2	1.1	12	14 14								¥.	- 3
	ú.	- 6	- 6	- 61	- 61	6	- 6	10	10	1.6	- 16	16 16								15	10
	6	- 10	12	16	16	l ő	ő	10	10	10	ii -	22 13						1		8	8,
	2	- 2	- 2	- 2	2	ö	ö	0	0	10	10	10 10							1	N.	- 52
i -	ō	0	0	ō	0	0	ō	- 2	2	1	1	2 2						. 1		х.	- 3
	13	15	15	15	15	7	7	15	15	15	18	18 19	Co	ernes				н з		18	8
	1	1	1	1	1	1	1	1	1	2	2	2 2	date	and so a COO						13	1.4
	25	25	26	- 25	26	22	22	- 24	- 24	34	34	35 34	aha	energ cau) OB atrios	(10)		6 3	ő –	÷.	- 12
	47	-47	- 47	-45	-47	-42	42	-42	- 42	135	135	135 135	e les	ators ope	on-sinps	(20)		0 2		12	- 12
•	20	20	20	19	20	16	16	16	16	20	20	20 20	il con	dile cot	Longing	1000		4 1	(10	- 12
	2	1	- 2	- 2		2		- 2	- 2	2	2	1.1	free	cell (ND	a sardar	(20)		2 4	£	ñ.	- 47
	-	- 10	- 12	- 48	48	42	42	- 40	- 40	48	48	47 48		-cent 1.4-ats	rine (20)		- 1	5 1	ξ	19	19
	2	- 7	- 7	- 1	- 7			- 1	- 1	8		8 8	geu	color (165	nps (an)			ő 1	έ.		
	-					1	- 6	- 3		1	1	2.1		conc (15)	,			8 2		28	27
	ú.	- 6	- 44	- 44		6	- 6	- 6		1.1				and the second	at Linearia	~ 1360			÷ .	TT -	- 15
	÷.	13	- 12	- 12	- 2	2	- 2	- 12		16	16	16 1		mainter f	R. string	MD		7 1		16	19
	÷.	8	- 9	- <u>é</u>		6	- 6	- 6	- 7	12	12	12 12	040	printer of	otHistri	05 (20)		i i	i i	12	15
-	61	67	67	67	- 67	67	67	-67	67	87	87	87 B	nin	saveld.e	of and any	1500		5 5	5	55	23
2	96	296	300	310	313	242	242	285	286	4.99	460	473 464	100	alver-0	8-string (300		i i	3	10	10
-	-					-	-	-						alyzer.ce	ot 11-strip	06 (20)		0 1	ñ	÷.	7
													100	in ont18	strins (2)	11		ă î	2	15	- 13
													tetr	s-cot14-s	strips (17	ĩ		i i	D.	π.	- íí
													trat	aport-opt	08-strips	000		4 i	4	13	- 13
													trat	non-none	11-strips	(20)	. i	0 1	Ó.	9	8
													tran	sport-opt	14-strips	(20)		8	Ň	8	- 7
													1010	dworking	2-00108-1	strips ()	m :	10 2	ō	19	26
													WO	dworkin	e-opt11-	strips (2	m 1	4 1	ă –	iă-	18
					overa	22		- 1	arb	inv	rnd	GZD	200	otravel (2	00			a i	š –	1	7
				1.5	Innet	1500			28	27	21	1 29	-				101			-	1014
				1.5	an p-Office							1 47	No.	n (m 27)			100	- IM			1116
					(leichen)	15		- 1	28	27	28	28									
				1	llocks	35)	00		28	27	28	28	-12	12	12	12	12.1				14
				1	llocks htaNe	(35) twork	(20)		28 12	27	28 12	28 13	-12	12	12	12	12		131		14
					llocks lotaNe lepot ((35) twork 22)	(20)		28 12 7	27 12 7	28 12 7	28 13 7	12	12	12	12	12		13		14
					Hocks MataNe Nepot (NriverL	(35) twoels 22) og (2	: (20) 9)		28 12 7 13	27 12 7 14	28 12 7 13 20	28 13 7 14 22	12 7 13	12 7 13	12 7 13	12 7 13	12 7 13		13 10 13		14 13 15
					llocks (htaNe hepot (hevator llevator	(35) twork 22) og (3 rs (30)	; (20) 9)		28 12 7 13 22	27 12 7 14 22	28 12 7 13 20	28 13 7 14 22 24	12 7 13 22 16	12 7 13 22	12 7 13 22	12 7 13 22	12 7 13 22 21		13 10 13 22 11		14 13 15 24
					llocks (httaNe hepot () hevator recoell	(35) twoels 22) og (30 rs (30) (80)	: (20) 9)		28 12 7 13 22 15 2	27 12 7 14 22 15	28 12 7 13 20 15	28 13 7 14 22 24	12 7 13 22 16	12 7 13 22 12	12 7 13 22 15	12 7 13 22 15	12 7 13 22 21		13 10 13 22 33		14 13 15 24 64
					llocks) hataNe hepot () hevator reccell irid (5) fikian	(35) twork 22) og (2) ts (30) (80)	: (20) 0)		28 12 7 13 22 15 2 10	27 12 7 14 22 15 2	28 12 7 13 20 15 1	28 13 7 14 22 24 2 10	12 7 13 22 16 2	12 7 13 22 12 2	12 7 13 22 15 2	12 7 13 22 15 2	12 7 13 22 21 2 10		13 10 13 22 33 2		14 13 15 24 64 3
					llocks (bataNe Nepot (briverL levator freecell frid (5) liking ((35) twork 22) og (2 s (30) (80) (20)	: (20) 0)		28 12 7 13 22 15 2 10 27	27 12 7 14 22 15 2 10	28 12 7 13 20 15 1 9 25	28 13 7 14 22 24 2 10	12 7 13 22 16 2 9 25	12 7 13 22 12 2 8	12 7 13 22 15 2 9	12 7 13 22 15 2 9	12 7 13 22 21 2 10 25		13 10 13 22 33 2 9		14 13 15 24 64 3 14
					llocks hataNe hepot () hevator reccell fitid (5) liking (ogistic	(35) twork 22) og (2 cs (30) (80) (80) (20) s (63)	: (20) 0)		28 12 7 13 22 15 2 10 27	27 12 7 14 22 15 2 10 27	28 12 7 13 20 15 1 9 25	28 13 7 14 22 24 2 10 27	12 7 13 22 16 2 9 25	12 7 13 22 12 2 8 25	12 7 13 22 15 2 9 25	12 7 13 22 15 2 9 25	12 7 13 22 21 2 10 25		13 10 13 22 33 2 9 27		14 13 15 24 64 3 14 34
				E L L E F C F L M	lleeks) bataNe bepot C briverL llevator freecell fiking (ogistic fprime	(35) twork 22) og (2) (80) (80) (20) s (63) (35)	(20) 0)		28 12 7 13 22 15 2 10 27 23	27 12 7 14 22 15 2 10 27 25	28 12 7 13 20 15 1 9 25 22	28 13 7 14 22 24 2 10 27 23 23	12 7 13 22 16 2 9 25 23	12 7 13 22 12 2 8 25 23	12 7 13 22 15 2 9 25 22	12 7 13 22 15 2 9 25 22	12 7 13 22 21 2 10 25 25		13 10 13 22 33 2 9 27 24		14 13 15 24 64 3 14 34 31
				E L L L E F C F L N N	llecks) httaNe hepot (hevator hevator reccell irid (5) liking (ogistic fprime fystery	(35) twork 22) og (2) (80) (80) (20) s (63) (35) (19)	: (20) 9)		28 12 7 13 22 15 2 10 27 23 16	27 12 7 14 22 15 2 10 27 25 17	28 12 7 13 20 15 1 9 25 22 15	28 13 7 14 22 24 2 10 27 23 17	12 7 13 22 16 2 9 25 23 17	12 7 13 22 12 2 8 25 23 17	12 7 13 22 15 2 9 25 22 17	12 7 13 22 15 2 9 25 22 17	12 7 13 22 21 2 10 25 25 17		13 10 13 22 33 2 9 27 24 17		14 13 15 24 64 3 14 34 31 19
				PELEPOPENNN	Ilocks (httaNe hepot C hriverL levator roccell irid (5) liking (ogistic fprime fystery iomyst	(35) twoek 22) og (20 (80) (20) s (63) (35) (19) ery G	: (20) 0) 0		28 12 7 13 22 15 2 10 27 23 16 16 21	27 12 7 14 22 15 2 10 27 25 17 16	28 12 7 13 20 15 1 9 25 22 15 14 14	28 13 7 14 22 24 20 27 23 17 17	12 7 13 22 16 2 9 25 23 17 15	12 7 13 22 12 2 8 25 23 17 14	12 7 13 22 15 2 9 25 22 17 18	12 7 13 22 15 2 9 25 22 17 18	12 7 13 22 21 22 10 25 25 17 16		13 10 13 22 33 2 9 27 24 17 18		14 13 15 24 64 3 14 34 31 19 20
				PETEFOLINNNO	Ilocks (httaNe hepot C hriverL llevator receell irid (5) liking (ogistic fprime fystery fomyst ppensta	(35) twork (22) og (30) (80) (80) (20) s (63) (15) (15) ery (5) cks (15)	(20) () () () () () () () () () () () () ()		28 12 7 13 22 15 2 10 27 23 16 16 31	27 12 7 14 22 15 2 10 27 25 17 16 31	28 12 7 13 20 15 1 9 25 22 15 14 31	28 13 7 14 22 24 2 10 27 23 17 17 17 13	12 7 13 22 16 2 9 25 23 17 15 31	12 7 13 22 12 2 8 25 23 17 14 30	12 7 13 22 15 2 9 25 22 17 18 31	12 7 13 22 15 2 9 25 22 17 18 31	12 7 13 22 21 22 10 25 25 17 16 31		13 10 13 22 33 2 9 27 24 17 18 31		14 13 15 24 64 3 14 34 31 19 20 34
				PEEEFOFLANNOOD	Ilocks (httaNe hepot (hriverL levator receell irid (5) liking (ogistic fprime fystery fornyst pensta hrgSyn	(35) twork 22) og (2 s (30) (80) (20) s (63) (35) (19) ery (5 cks (th-spl	: (20) 0) 0 0 0 0 0 0 0 0 1 1 (20	5	28 12 7 13 22 15 2 10 27 23 16 16 31 5 9	27 12 7 4 22 15 2 10 27 25 17 16 31 5 31 5	28 12 7 13 20 15 1 9 25 22 15 14 31 14	28 13 7 14 22 24 2 10 27 23 17 17 31 15 30	12 7 13 22 16 2 9 25 23 17 15 31 14	12 7 13 22 12 2 8 25 23 17 14 30 10	12 7 13 22 15 2 9 25 22 17 18 31 15 2	12 7 13 22 15 2 9 25 22 17 18 31 52	12 7 13 22 21 2 10 25 25 17 16 31 15		13 10 13 22 33 2 9 27 24 17 18 31 15 20		14 13 15 24 64 3 14 34 31 19 20 34 10
				E E E E F C F L M M NO O P	Ilocks (httaNe hepot (hriverl, levator receell irid (5) liking (orgistic fprime fystery iomyst pensta hygSyn arcprin	(35) twork 22) og (2 (80) (80) (30) (30) (30) (30) (30) (30) (30) (3	: (20) 0) 0 0 0 0 0 90) 11 (20 0)	ŋ	28 12 7 13 22 15 2 10 27 23 16 16 31 5 9 9 9 9 9 9 9 10 10 10 10 10 10 10 10 10 10	27 12 7 4 22 15 20 27 25 10 27 25 17 16 31 52 20	28 12 7 13 20 15 1 9 25 22 15 14 31 14 19	28 13 7 14 22 24 10 27 23 17 31 15 22 2	12 7 13 22 16 2 9 25 23 17 15 31 14 19	12 7 13 22 12 2 8 25 23 17 14 30 10 19	12 7 13 22 15 2 9 25 22 17 18 31 15 22	12 7 13 22 15 2 9 25 22 17 18 31 5 22	12 7 13 22 21 2 10 25 25 17 16 31 15 18		13 10 13 22 33 2 9 27 24 17 18 31 15 20		14 13 15 24 64 3 14 34 31 19 20 34 10 30
				E E E E E E E E E E E E E E E E E E E	Ilocks i hataNe hepot () hriverL llevator reecell irid (5) liking i ogistic fprime fystery formyst prensta hypensta hypensta hypensta	(35) twork 22) og (2 s (30) (80) (20) s (63) (19) ery G cks (th-spl tter ((40)	: (20))) (1) (20) (20) (20) (20) (20) (20) (20) (20	ŋ	28 12 7 13 22 15 2 10 27 23 16 16 31 5 9 9	27 12 7 14 22 15 20 27 25 17 31 22 9 14 25 17 31 5 22 9 14	28 12 7 13 20 15 1 9 25 22 15 14 31 14 19 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	28 13 7 14 22 10 27 23 17 17 31 15 22 9 7	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10	12 7 13 22 15 2 9 25 22 2 9 25 22 17 18 31 15 22 212	12 7 13 22 15 2 9 25 22 17 18 31 15 22 17	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 8		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13		14 13 15 24 64 3 14 34 31 19 20 34 10 30 31 37
				E E F F C F F L M N N C C P P P	Ilocks (httaNe hepot (hriverL levator reecell irid (5) liking (ogistic fprime fystery fomyst hpensta hgSyn areprin arking legsol ((35) twoek (22) og (3 x (30) (30) (30) (35) (19) ery G ccks ((40) h-spl ter (7 (40) 36)	(20)) (0) (0) (0) (1) (2) (2) (2) (2) (2) (2) (2) (2	ŋ	28 12 7 13 22 15 2 10 27 23 16 16 31 5 19 9 35	27 12 7 14 22 15 20 27 25 17 31 5 22 9 34	28 12 7 13 20 15 1 9 25 222 15 14 31 14 19 6 33	28 13 7 14 22 24 2 10 27 23 17 17 31 5 22 9 35 35	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10 34	12 7 13 22 15 2 9 25 22 17 18 31 15 22 12 35	12 7 13 22 15 2 9 25 22 17 18 31 15 22 12 34	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 34		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35		14 13 15 24 64 3 14 34 31 19 20 34 10 30 13 35
				E E F C F F C F F F C F F F F F F F F F	llocks i hataNe hepot (hriverL llevator receell irid (5) liking i orgistic fiprime fystery formyst hypensta hypensta hypensta hypensta hypensta hypensta hypensta	(35) tweek (22) (30) (30) (30) (30) (30) (30) (30) (31) (31) (31) (31) (32) (31) (32) (32) (32) (32) (33) (33) (33) (33	(20) (0) (0) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	ŋ	28 12 7 13 22 15 2 10 27 3 16 16 31 15 19 9 35 18 19 18 18 18 19 18 18 18 19 18 18 18 18 18 18 18 18 18 18 18 18 18	27 12 7 14 22 15 2 10 27 25 17 16 31 5 2 9 34 18 2 9 34	28 12 7 13 205 15 222 15 14 31 4 19 6 33 17 25	28 13 7 14 22 24 20 27 210 27 213 17 31 15 29 35 18 20 17 18 20 27 31 18 20 20 27 31 18 20 20 20 20 20 20 20 20 20 20	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10 34 17	12 7 13 22 9 25 22 17 18 31 15 22 17 18 31 5 22 12 35 18	12 7 13 22 15 2 9 25 22 17 18 31 15 22 12 34 18	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 34 17		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18		14 13 15 24 64 3 14 34 31 19 20 34 10 30 13 35 25
				E E E E E E E E E E E E E E E E E E E	Ilocks hataNe hepot (hriverl, levator receell irid (5) liking (ogistic fprime fyster) iomyst hystery iomyst hystery iomyst hypersta hypersta hypersta hypersta hypersta hypersta hypersta hypersta	(35) tweek (22) (30) (30) (30) (30) (30) (30) (30) (31) (31) (31) (31) (31) (31) (31) (31	(20) (0) (0) (0) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	ŋ	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 0	27 12 7 14 22 15 2 10 27 5 17 16 31 15 22 9 34 18 12 2 9 34	28 12 7 13 205 15 22 22 15 14 31 14 9 33 17 10 9	28 13 7 14 22 24 20 27 23 17 17 15 22 9 35 18 12 2	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 11	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10 34 17 9	12 7 13 22 15 2 9 25 22 17 18 31 15 22 12 35 18 12 18 12 22 12 35	12 7 13 22 9 25 22 17 18 31 15 29 25 22 17 18 31 15 21 2 17 18 31 15 21 2 9 25 22 21 2 17 18 31 22 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 21 2 10 25 25 17 16 31 15 8 8 34 17 12		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 2		14 13 15 24 64 3 14 34 31 19 20 34 10 30 13 35 25 18
				E E E E E E E E E E E E E E E E E E E	Ilocks httaNe httpot (httpot	(35) twork (22) (30) (30) (30) (30) (30) (35) (19) (19) (19) (19) (19) (19) (19) (19	(20) (20) (30) (30) (5)) (5))	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9	27 12 7 14 22 15 2 10 27 5 17 16 31 15 22 9 34 18 22 9 34	28 12 7 13 20 15 15 22 25 225 14 31 4 19 6 33 17 10 7 9	28 13 7 14 22 24 2 10 27 23 17 31 15 22 9 35 18 18 9 9 35 18 19 19 10 27 21 21 21 21 21 21 21 21 21 21	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 11 9	12 7 13 22 12 2 8 25 23 17 14 30 10 10 34 17 9 9 9	12 7 13 22 15 2 9 25 22 17 18 31 15 22 12 12 12 12 15 22 12 15 22 17 18 31 15 22 25 15 22 15 29 25 15 22 15 22 15 29 25 15 22 29 25 22 15 22 29 25 22 15 22 22 25 22 25 22 22 25 22 25 22 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 25 22 22	12 7 13 22 9 25 22 17 18 31 15 29 25 217 18 31 15 212 34 18 12 9 9	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 34 17 12 9		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9		14 13 15 24 64 31 14 34 31 19 20 34 10 30 13 35 25 18 0 0
				E E E E E E E E E E E E E E E E E E E	Ilocks hataNe hepot (hepot (hevator levator levator levator liking (opjistic dprime dystery formyst hystery	(35) twork (22) og (3 (30) (30) (30) (35) (19) ery (7 (40) (40) (40) (36) ttank (5) grms (40)	(20))))))))))))))	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 1	27 12 7 14 22 15 2 10 27 25 17 16 31 15 22 9 34 18 12 9 11 12 15 2 10 27 12 15 2 10 27 12 15 2 10 27 12 12 12 12 15 2 10 27 12 12 12 12 12 12 12 12 12 12 12 12 12	28 12 7 13 20 15 1 9 22 22 15 14 31 19 6 33 170 7 9	28 13 7 14 22 24 20 27 23 17 31 15 29 35 18 19 9 9	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 11 9 9	12 7 13 22 12 2 8 25 23 17 14 30 10 10 10 34 17 9 9 9 9	12 7 13 22 15 2 9 25 22 17 18 31 15 22 17 18 31 15 22 12 35 18 12 9 9 9 9	12 7 13 22 15 2 9 25 22 17 18 31 15 22 17 18 31 5 22 12 34 18 22 9 9 9 9 9	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 34 17 12 9 9		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9 9		14 13 15 24 64 3 14 34 14 31 19 20 34 10 30 13 35 25 18 0 9
				E E E E E E E E E E E E E E E E E E E	Hocks of the second sec	(35) twork (22) (30) (30) (30) (30) (35) (19) (35) (19) (35) (19) (40) (36) (40) (40) (40) (36) (40) (36)	(20) (50) (50) (50) (50) (50) (51) (51) (51)	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8	27 12 7 14 22 5 2 10 27 25 17 16 31 15 22 9 34 18 12 9 34 18 12 9 11 27 12 15 2 10 27 25 17 16 31 11 27 25 17 16 31 11 27 16 27 17 16 27 17 17 16 27 17 17 16 17 17 16 17 17 16 17 17 17 17 17 17 17 17 17 17 17 17 17	28 12 7 13 20 15 1 9 25 22 15 14 31 14 9 6 33 17 10 7 9 7	28 13 7 14 22 24 2 10 27 21 17 15 22 9 5 5 18 12 9 8 8 12 12 12 12 12 12 12 12 12 12	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 11 9 9 14	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10 34 17 9 9 9 13	12 7 13 22 9 25 22 17 18 31 15 22 12 35 18 12 9 9 9 15	12 13 22 15 29 25 22 17 18 31 15 22 12 15 29 25 22 17 18 31 15 29 20 20 20 20 20 20 20 20 20 20	12 7 13 22 21 10 25 25 17 16 31 15 18 8 34 17 12 9 9 10		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14		14 13 15 24 64 3 14 34 31 19 20 30 30 30 30 30 30 30 30 30 30 9 8 8
				E E E E E E E E E E E E E E E E E E E	Hocks i hutabe hepot (briverL, briverL, liking or ogistic dystery formyst dyst dystery formyst dystery formyst dystery formyst dystery formyst dyst dyst dyst dyst dyst dyst dyst d	(35) twork (22) (30) (30) (30) (30) (30) (30) (30) (30	(20) (30) (30) (30) (5)) (5))	77)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 4	27 12 7 14 22 15 2 10 27 5 17 6 31 31 5 22 9 34 8 18 2 9 34 18 12 9 12 15 15 15 15 15 15 15 15 15 15 15 15 15	28 12 7 3 20 15 1 9 25 22 15 14 14 19 6 33 17 10 7 9 7 16	28 13 14 22 24 20 22 10 27 23 17 17 15 29 55 18 29 9 8 18 20 4 20 17 17 17 15 29 55 18 20 18 20 18 20 18 20 17 17 17 15 20 18 20 18 20 18 20 17 17 17 18 18 20 18 20 18 18 18 18 18 18 18 18 18 18	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 11 9 9 14 15	12 7 13 22 2 8 25 23 17 14 30 10 10 10 34 17 9 9 9 9 13 14	12 7 13 22 15 2 9 25 217 18 31 15 22 17 18 31 5 12 9 9 9 15 16	12 13 22 15 2 25 22 17 18 31 15 22 17 18 31 15 22 15 29 25 22 17 18 31 15 29 25 20 15 29 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 25 20 20 25 20 20 25 20 20 20 20 20 20 20 20 20 20	12 7 13 22 21 20 25 25 17 16 31 15 18 8 34 17 12 9 9 10 16		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 2		14 13 15 24 64 3 14 31 19 20 34 31 19 20 34 10 30 30 30 35 25 18 8 0 9 9 8 8 18
				E E E E E E E E E E E E E E E E E E E	Hocks i hataNe bepot (briverL llevator priverL llevator persta hy	(35) twork (22) (30) (30) (30) (30) (30) (30) (30) (30	(20) (3) (30) (50) (50) (50) (50) (11 (2) (2) (3) (3) (3) (3) (3) (3) (3) (3	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6	27 12 7 14 22 5 2 10 27 25 17 16 31 18 22 9 34 18 12 9 34 18 12 9 11 12 15 2 17 16 11 27 17 17 16 11 27 17 17 18 19 27 19 27 19 27 19 27 19 27 19 27 19 27 19 27 19 27 19 27 19 19 19 19 19 19 19 19 19 19 19 19 19	28 12 20 15 1 1 9 9 25 22 15 14 14 19 6 33 3 17 7 9 7 7 9 7 7 16 4 4	288 13 7 14 22 24 2 10 27 23 17 17 15 22 9 35 18 12 9 9 8 16 6 16 17 17 18 18 18 18 18 19 19 19 19 19 19 19 19 19 19	12 7 13 22 16 2 9 25 23 17 15 31 14 19 10 34 17 15 31 14 19 9 9 4 15 4	12 7 13 22 12 2 8 25 23 17 14 30 10 19 10 34 17 9 9 9 9 13 14 4 4	12 7 13 22 9 25 22 17 18 31 15 22 12 12 35 18 12 9 9 15 16 6	12 13 22 15 29 25 227 18 31 15 22 12 34 18 12 29 9 9 15 12 15 29 25 217 18 31 15 29 15 29 25 217 18 31 15 29 25 217 18 31 15 29 25 217 18 31 15 29 25 217 18 31 15 29 25 217 18 31 15 29 25 217 18 31 15 212 31 15 212 15 217 18 31 15 212 15 217 18 18 18 18 19 19 10 10 10 10 10 10 10 10 10 10	12 7 13 22 21 22 10 25 25 25 17 16 31 15 8 34 17 12 9 9 10 16 6		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7		14 13 15 24 64 3 14 34 14 34 19 20 34 10 30 30 35 25 18 8 0 9 9 8 18 13
				E E E E E E E E E E E E E E E E E E E	Hocks - hatabe hepot (hervato hervata hervato	(35) twork (22) (30) (30) (30) (30) (30) (30) (30) (35) (19) (19) (10) (40) (36) (40) (40) (36) (40) (36) (40) (36) (40) (36) (40) (36) (40) (36) (40) (37) (40) (40) (37) (40) (40) (37) (40) (40) (40) (40) (40) (40) (40) (40	(20))) (0) (50) (50) (50) (1) (1) (1) (1) (1) (1) (2) (2) (2) (2) (2) (2) (2) (2	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 39 1	27 12 7 14 215 2 10 27 5 217 16 31 15 22 9 34 18 22 9 34 18 12 9 11 12 6 6 29 21 2 5 5 17 16 31 18 22 9 34 18 29 18 20 19 19 19 19 10 10 10 10 10 10 10 10 10 10 10 10 10	28 12 7 13 20 15 15 15 15 15 22 25 22 25 25 22 25 15 14 31 14 19 6 33 17 10 0 7 7 9 7 7 6 33 33 17 0 9 9 25 25 25 25 25 25 25 25 25 25 25 25 25	28 13 14 224 20 10 27 23 17 15 22 9 35 18 22 9 8 16 6 39 30 30 30 30 30 30 30 30 30 30	12 7 13 22 16 2 9 25 23 17 15 31 14 19 9 4 15 31 14 9 9 14 1 5 4 30 9	12 7 13 22 2 8 25 23 17 14 30 10 10 34 17 9 9 9 9 13 14 30 10 34 4 30	12 7 13 22 9 25 22 17 18 31 15 22 22 17 18 31 15 22 12 35 18 31 15 29 9 15 16 6 30	12 7 13 22 9 9 25 22 22 17 18 31 15 22 22 17 18 31 15 22 22 17 18 31 15 16 16 17 18 31 15 16 17 18 18 31 18 19 25 25 29 9 25 20 20 25 20 20 25 20 20 25 20 20 20 20 20 20 20 20 20 20 20 20 20	12 7 13 22 21 2 10 25 25 17 16 31 15 8 8 34 17 2 9 9 10 6 6 30		13 10 13 22 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7 30		144 13 15 244 644 31 14 34 31 19 200 34 31 19 200 34 30 30 9 8 8 18 18 13 330
				E E E E E E E E E E E E E E E E E E E	Hocks i hataNe bepot (2 hivertal lievator orgistic fiprine ybystery by	(35) twork (22) (30) (30) (30) (30) (30) (35) (35) (35) (35) (35) (35) (35) (35	(20))) (0) (50) (77)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 30 11	27 12 7 4 22 25 2 10 27 25 17 16 31 5 22 9 34 8 12 9 34 8 12 9 11 12 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 12 12 9 34 12 12 9 34 12 12 9 34 12 12 9 34 12 12 12 12 12 12 12 12 12 12 12 12 12	28 12 7 13 20 15 15 15 14 31 19 6 33 31 10 7 7 7 16 4 30 9 9 9 7 7 8 9 9 9 9 9 9 9 9 9 9 15 15 15 15 15 15 15 15 15 15 15 15 15	288 13 7 14 224 20 10 27 23 17 31 5 22 9 35 18 12 9 8 16 6 30 12 12 12 12 12 12 12 12 12 12	12 7 7 13 22 2 9 25 21 3 1 2 9 25 21 7 15 31 4 19 10 3 4 19 10 3 4 17 19 9 9 25 3 11 11 11 11 11 11 11 11 11 11 11 11 1	12 7 13 22 2 2 2 2 2 2 2 2 2 2 2 3 12 2 2 2	12 7 13 22 9 25 22 17 18 31 15 22 22 17 18 31 15 21 2 35 18 31 15 12 9 9 9 15 16 6 30 11	12 7 13 22 9 25 22 22 9 25 22 22 17 18 31 5 22 12 34 18 12 9 9 15 16 6 6 6 0 0 11	12 7 13 22 21 10 25 25 17 16 31 5 18 8 34 17 12 9 9 10 16 6 6 0 10		13 10 13 22 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7 30 12		144 133 15 244 644 33 144 341 19 200 344 100 303 303 30 9 8 8 183 300 15 5
				E E E E E E E E E E E E E E E E E E E	Höcks i hataNe bepot (briverL llevator irrecell liking gogistic dprime gogistic dprime gogistic dprime gogistic dprime provest https:// taking provest https:// taking ta	(35) (work (22) (30) (30) (30) (30) (30) (30) (30) (30	(20))) (50) (5))	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 39 11 7	27 12 7 4 22 15 2 10 27 16 31 15 2 9 34 18 12 9 34 18 12 9 34 18 12 9 34 18 12 9 34 18 12 9 34 18 16 6 9 11 2 15 16 16 16 16 16 16 16 16 16 16 16 16 16	28 12 7 3 20 15 1 9 25 22 2 15 1 4 31 31 14 14 14 19 6 33 3 17 7 9 7 7 6 4 4 30 9 7 8 6 6 6 6 6 6 6 7 8 9 7 7 13 13 20 0 15 15 15 15 15 15 15 15 15 15 15 15 15	288 13 7 14 224 20 10 27 23 17 31 15 22 9 35 18 12 9 9 8 16 6 30 12 7 7 7 7 7 7 7 7 7 7 7 7 7	12 7 13 222 16 2 9 25 31 15 31 14 9 9 4 15 31 11 9 9 4 4 30 11 6	12 7 13 22 2 8 25 23 17 14 30 10 10 10 10 10 34 17 9 9 9 13 14 4 30 9 5 6	12 7 13 22 9 9 25 22 22 17 18 31 12 22 12 18 31 12 22 12 18 31 12 29 9 9 15 6 30 11 7	12 7 13 22 9 9 25 22 22 17 18 31 15 22 12 12 12 12 13 14 18 12 9 9 9 15 6 6 30 0 11 7 7	12 7 13 22 10 25 17 16 31 15 18 8 34 17 12 9 9 9 10 6 30 16 6 30 16 6		13 10 13 22 33 2 9 27 24 17 18 31 20 13 35 18 20 9 14 16 7 30 12 7		144 13 15 244 64 34 31 19 200 30 30 30 30 30 30 9 9 8 18 13 30 5 15 13
				E E E E E E E E E E E E E E E E E E E	Hocks shataNe hepot (2) hypert (2) hypert (2) hypert (2) hyperstation ((35) twork (22) (30) (30) (30) (30) (30) (30) (30) (30	(20)) (0) (0) (50) (50) (50) (1) (50) (1) (50) (1) (1) (1) (1) (1) (1) (1) (1	73	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 39 11 7 23	27 12 7 4 22 15 2 10 27 25 17 6 3 15 22 9 34 18 12 9 34 18 22 9 18 22 9 18 18 22 9 18 18 22 9 18 18 22 9 18 18 18 18 18 18 18 18 18 18 18 18 18	28 12 7 13 20 15 14 14 19 6 33 34 17 10 7 7 16 30 9 7 16 31 19 9 255 225 255 255 255 15 14 14 19 9 7 15 16 17 17 19 9 255 255 255 15 16 17 17 17 17 17 17 17 17 17 17	28 13 7 14 224 20 27 23 17 15 22 9 35 18 29 9 8 16 6 30 12 7 23 16 23 23 17 18 22 9 8 18 22 9 8 18 22 23 23 17 18 22 23 17 18 22 23 17 18 22 23 17 18 22 23 17 18 22 23 17 18 22 23 17 18 22 9 8 18 22 9 9 8 18 18 22 9 9 8 18 18 18 18 18 18 18 18 18	12 7 13 22 16 2 9 25 23 17 15 11 14 19 10 34 17 11 9 9 14 15 16 2 9 17 15 11 14 19 10 10 10 10 10 10 10 10 10 10	12 7 13 22 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 9 25 22 22 17 18 31 15 22 9 9 25 22 217 18 31 15 22 17 18 31 15 22 9 9 25 22 217 18 31 15 22 9 9 25 22 217 18 31 15 22 22 25 22 217 18 31 15 22 22 25 22 217 18 31 22 22 25 22 22 25 22 22 25 22 22 25 22 22	12 7 13 22 9 9 25 22 9 9 25 22 17 18 18 31 15 22 12 17 18 31 15 22 12 9 9 9 15 16 6 30 17 7 22 22 22 22 22 22 22 22 22 22 22 22	12 7 13 22 21 2 10 25 25 17 16 31 15 18 8 34 17 12 9 9 10 6 30 10 6 30 10 6 22		13 10 13 22 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7 30 12 7 23		144 13 15 244 64 31 14 31 19 200 30 30 13 35 52 5 18 0 9 9 8 18 13 30 15 13 32 22
				E E E E E E E E E E E E E E E E E E E	Hocks is wata Neepot (briver). Jevatos Pepot (briver). Jevatos processi (briver). Jevatos processi (brid) (bri	(35) tweek (22) (30) (30) (30) (30) (30) (30) (31) (31) (31) (32) (40) (40) (36) (40) (36) (40) (36) (36) (36) (36) (36) (36) (36) (36	(20)) (0) (0) (1) (1) (1) (1) (1) (1) (1) (1	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 39 11 7 23 16	27 12 7 4 22 15 2 10 27 15 2 10 27 15 2 10 27 16 3 15 22 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 34 8 12 9 12 12 12 9 12 12 12 12 12 12 12 12 12 12 12 12 12	28 12 13 20 15 14 19 9 25 22 15 14 19 9 25 15 14 19 9 7 7 9 7 15 13 14 19 9 7 15 15 15 15 15 15 15 15 15 15	28 13 7 14 224 20 27 23 17 17 15 29 35 18 29 9 8 16 6 30 12 7 23 17 17 17 18 29 9 8 16 6 30 12 7 23 17 17 23 17 17 17 17 17 17 17 17 17 17	12 7 13 22 16 2 9 22 23 17 15 2 9 9 14 19 9 14 19 9 14 10 34 11 9 9 14 10 34 11 15 40 10 10 10 10 10 10 10 10 10 1	12 13 22 12 2 2 8 25 23 17 14 20 10 10 10 10 10 34 10 10 34 10 10 34 17 12 2 8 25 23 17 14 4 30 9 9 9 13 14 4 30 25 25 23 17 12 25 23 17 17 25 25 25 25 23 17 17 25 25 25 25 25 25 25 25 25 25 25 25 25	12 7 13 22 15 2 9 25 22 7 18 31 5 22 2 5 22 7 18 31 5 22 2 5 22 7 18 31 5 29 9 5 5 22 7 7 18 31 5 29 9 5 5 20 7 7 18 8 3 15 5 29 9 5 5 20 7 7 18 8 5 18 5 18 5 18 5 18 5 18 5 18	12 7 13 22 9 9 22 22 17 31 15 22 22 17 18 31 15 22 23 4 18 21 2 34 18 21 2 34 18 21 2 34 17 7 21 2 34 18 2 34 31 31 31 22 34 31 31 31 31 31 31 31 31 31 31 31 31 31	12 7 13 22 10 25 17 16 31 25 17 16 31 31 15 8 34 17 29 9 9 10 16 6 6 9 10 10 5 22 11 21 12 21 12 25 17 16 31 25 17 16 25 17 16 25 17 20 10 20 17 20 17 20 10 20 17 20 10 20 20 10 20 10 20 20 20 10 20 20 20 20 20 20 20 20 20 20 20 20 20		13 10 13 22 33 2 9 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7 30 12 7 23 36		144 13 15 244 644 34 34 34 14 34 10 30 30 35 25 25 18 8 0 9 9 8 8 18 30 15 13 22 23 0
				E E E E E E E E E E E E E E E E E E E	Hocks is hataNe hepot (briver). liking protect liking prime prim prime prime prime prim prime prime	(35) tweek (22) og (3 (3 (6 (6 (7 (7 (7 (7 (7 (7 (7 (7 (7 (7 (7 (7 (7	(20)) (0) (0) (0) (50) (50) (50) (1) (50) (30) (50) (30) (50) (30) (50) (30) (50) (30) (, 2) 3)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 30 11 7 23 16 19	27 12 7 4 22 5 2 10 27 25 7 16 31 8 22 9 34 8 12 9 34 8 12 9 11 16 6 29 11 6 29 11 20 16 20 17 16 20 17 16 20 20 20 20 20 20 20 20 20 20 20 20 20	28 12 20 15 22 22 22 22 22 22 22 25 15 14 19 9 25 15 15 14 19 9 7 6 4 30 9 9 6 20 0 7 9 7 6 20 0 15 15 15 15 15 15 15 15 15 15 15 15 15	28 13 7 14 22 24 24 24 27 23 17 17 17 15 22 9 9 35 18 2 9 9 8 8 16 6 30 12 7 7 7 31 15 22 15 15 22 15 15 15 15 15 15 15 15 15 15	12 7 13 22 9 23 17 15 14 19 9 14 19 9 14 15 4 30 16 6 20 5 22 23 17 15 14 19 10 10 10 10 10 10 10 10 10 10	12 7 13 22 2 8 25 23 17 14 30 10 10 10 10 10 10 34 17 9 9 9 9 13 14 4 30 9 9 9 5 23 23 23 24 25 25 25 25 25 25 25 25 25 25 25 25 25	12 7 13 22 9 25 22 17 18 15 22 9 25 22 17 18 15 22 12 15 2 9 25 22 17 18 15 22 15 23 15 15 23 15 15 23 15 15 23 15 15 23 15 15 23 15 15 15 15 15 15 15 15 15 15	12 7 13 22 9 9 25 22 22 17 18 31 31 22 22 22 17 18 31 31 22 22 23 4 8 12 29 9 9 5 5 22 25 22 21 5 15 5 29 9 9 25 22 22 15 5 15 20 9 25 22 22 22 22 22 22 22 22 22 22 22 22	12 7 13 22 2 10 25 25 17 16 31 15 8 34 17 12 9 9 10 6 6 30 10 6 22 14 22 22 12 23 24 25 25 17 16 31 25 25 25 17 26 25 26 25 26 27 20 25 26 27 20 20 20 20 20 20 20 20 20 20 20 20 20		13 10 13 22 23 29 27 24 17 18 31 15 20 13 35 18 12 9 9 14 16 7 30 12 7 23 36 20		144 13 15 244 64 31 34 34 34 30 30 30 30 30 30 9 9 8 8 18 13 30 9 9 8 13 30 222 30 30
				E E E E E E E E E E E E E E E E E E E	Hocks is watable wa	(35) (work (22) (35) (30) (30) (30) (30) (30) (30) (30) (33) (40) (40) (40) (40) (40) (36) (40) (36) (40) (36) (40) (36) (20) (20) (20) (20) (40) (40) (40) (20) (20) (20) (20) (20) (20) (20) (2	(20)) (30) (5)	20)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 35 18 12 9 9 8 16 6 39 11 7 23 16 19 13	27 12 27 14 22 25 2 10 27 25 17 16 31 5 2 2 9 34 8 12 9 34 8 12 9 11 12 25 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 27 5 2 10 2 10 2 10 2 10 2 10 2 10 2 10 2	28 12 13 20 15 14 19 22 22 15 14 19 22 22 15 15 16 30 10 7 7 16 30 9 9 6 20 15 15 15 15 15 15 15 15 15 15	28 13 7 14 22 14 22 10 27 17 17 17 17 17 17 17 18 22 9 9 8 16 6 30 9 9 8 16 10 17 17 17 17 17 17 17 17 17 17	12 7 13 22 9 22 23 17 13 14 19 9 14 15 4 30 11 6 20 36 20 12 14 15 14 10 10 14 10 10 10 10 10 10 10 10 10 10	12 12 7 13 22 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 15 2 9 25 22 17 18 15 2 9 25 22 17 18 15 2 9 25 22 17 18 15 2 9 25 22 17 18 15 2 9 25 22 17 18 15 2 9 25 22 12 15 2 9 25 22 12 15 2 12 2 17 18 18 18 12 2 9 9 16 6 10 17 17 18 18 12 2 12 2 12 2 12 2 12 2 12 2 12 2 16 17 17 18 18 12 2 12 2 3 18 12 2 12 2 3 18 12 2 12 2 3 18 12 2 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 9 9 25 22 22 17 18 31 15 22 22 17 18 31 15 22 22 17 18 31 15 22 29 9 15 16 6 6 30 11 7 22 31 5 22 23 23 5 22 23 24 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 25 5 22 2 5 5 22 2 5 5 22 2 5 5 22 2 5 5 22 2 5 5 22 2 2 5 5 2 2 2 5 5 2 2 2 2 5 5 2 2 2 2 2 2 5 5 2 2 2 2 2 2 2 5 5 2	12 7 13 22 10 25 25 17 16 31 15 8 34 7 10 25 25 17 6 31 15 8 34 7 10 25 25 17 6 31 15 8 34 7 10 25 21 2 10 25 25 17 20 10 25 25 17 20 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 17 10 25 25 10 25 25 10 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 10 25 25 10 25 25 10 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 10 25 25 25 25 25 25 25 25 25 25 25 25 25		13 10 13 22 23 29 27 417 18 31 15 20 13 31 15 20 13 31 16 7 30 12 7 23 30 20 12 23 30 20 27 23 30 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 27 20 20 20 20 20 20 20 20 20 20 20 20 20		14 13 15 24 64 64 3 14 31 9 20 30 30 30 35 25 25 18 8 0 9 8 8 13 30 15 13 22 230 30 30 31 3
				EDDDEFDEFDEFDEFDEFDEFDEFDEFDEFDEFDEFDEFD	Hocks is hataNee Peptot C hviverL. Alevator precedent irid (5) liking orgistic fiprime fystery formyst formyst fiprime fystery formyst fippenta tarking tarkin	(35) twock (22) (35) (30) (30) (30) (30) (30) (30) (30) (30) (30) (30) (30) (40) (40) (30) (40) (20) (40) (20) (40)	(20)) (30) (5)	7)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 9 9 8 16 30 11 7 23 16 31 15 9 9 8 16 30 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 31 17 23 16 23 16 23 16 23 17 23 16 23 16 23 17 23 16 23 16 23 17 23 17 23 16 23 17 17 17 17 17 17 17 17 17 17	27 12 12 14 22 15 210 27 25 16 31 15 29 34 18 29 34 18 29 34 18 29 34 18 29 34 18 29 10 29 10 20 20 20 20 20 20 20 20 20 2	288 12 7 7 13 200 15 1 1 9 9 25 22 25 22 25 25 22 25 25 22 25 15 1 1 1 9 9 25 22 25 22 25 15 1 1 1 1 9 9 25 22 25 25 22 25 25 25 25 25	288 13 7 144 2 244 2 10 27 27 10 27 17 17 17 17 17 17 17 17 17 1	12 7 13 22 23 16 2 9 9 25 23 17 15 31 14 19 10 10 13 14 19 10 10 10 10 10 10 10 10 10 10	12 12 7 13 22 12 2 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 9 25 22 2 9 25 22 2 17 18 3 15 22 9 9 5 15 16 6 30 11 7 22 36 30 11 7 22 5 31 3 15 2 2 9 9 35 31 3 15 2 2 9 2 5 5 2 2 2 5 5 2 2 2 5 5 2 2 2 5 5 2 2 5 5 2 2 2 5 5 3 15 2 2 9 2 5 5 2 2 5 5 2 2 2 5 5 2 2 2 5 5 2 2 2 5 5 2 2 2 5 5 5 2 2 2 5 5 5 2 2 2 5 5 5 5 2 2 2 5	12 7 13 22 2 9 9 25 22 22 17 7 18 31 15 22 22 22 12 34 18 31 12 22 23 4 18 18 31 17 7 22 2 9 9 9 15 16 5 20 9 25 22 22 22 23 22 24 22 25 22 22 25 22 22 22 22 23 22 22 23 22 22 23 22 22	12 7 13 221 2 10 25 25 25 17 16 31 15 18 8 34 17 12 9 9 10 6 6 30 6 6 30 16 6 31 0 6 22 12 9 10 12 12 25 25 17 16 31 25 25 25 17 16 31 25 25 25 25 25 25 25 25 25 25 25 25 25		13 10 13 22 9 27 23 33 2 9 27 23 13 29 27 23 15 20 13 35 18 12 9 9 14 16 7 20 30 12 20 27 27 23 31 29 27 27 24 17 18 31 20 20 27 27 27 27 27 20 27 20 27 20 27 20 27 20 27 20 27 20 20 20 20 20 20 20 20 20 20		14 13 15 24 64 3 14 4 31 19 20 0 30 30 30 30 30 30 30 30 30 30 30 30
				E E E F G F L L M M N O G P P P P P P B S S S S T T T V V Z Z G	Hocks is hataNee Peptot C herverL. Ilevation orgistic (5) liking orgistic flyrinee dystery flyrinee dystery	(35) twock (22) (35) (30) (30) (30) (30) (30) (30) (31) (35) (35) (35) (35) (35) (35) (35) (35	(20)) (30) it (20) (5)) (5))	50)	28 12 7 13 22 15 2 10 27 23 16 16 31 15 19 9 8 16 6 39 11 7 23 16 19 13 331 3331	277 127 144 222 152 200 277 255 217 16 31 18 22 9 34 18 12 9 11 16 6 29 215 215 215 215 215 215 215 215	28 12 7 13 20 15 1 9 9 25 22 25 26 27 27 26 27 27 27 27 27 27 27 27 27 27	288 13 7 14 22 24 10 27 27 23 17 17 17 17 17 17 17 17 18 18 18 18 18 18 18 18 18 19 9 9 9 9 9 9 9 9 9 9 9 9 9	12 7 13 22 9 9 25 23 17 15 31 14 19 10 34 17 11 19 9 9 14 15 30 34 17 11 10 34 20 34 20 34 17 17 31 10 34 10 34 20 35 23 31 17 17 31 22 23 31 17 18 25 25 23 31 17 18 20 25 25 23 31 17 18 20 25 25 23 31 17 18 20 25 25 23 31 17 18 20 25 25 23 31 17 18 20 25 25 25 23 31 17 18 19 25 25 25 23 31 17 18 19 25 25 25 25 23 31 17 17 18 19 25 25 25 25 25 25 20 25 25 20 25 20 25 25 20 25 20 25 20 20 20 20 20 20 20 20 20 20 20 20 20	12 7 33 22 2 2 8 8 25 23 23 17 14 30 10 19 10 34 4 4 4 4 9 9 5 5 36 15 36 15 36 22 22 331	12 7 13 2 9 9 25 2 2 2 2 2 2 2 2 2 2 2 2 2	12 7 13 22 9 9 25 22 17 18 31 15 22 22 17 18 31 15 22 22 17 18 31 15 22 22 17 18 31 15 22 22 17 18 31 15 22 22 9 9 25 5 22 22 17 7 18 31 15 22 22 9 9 25 5 22 22 17 7 18 31 15 22 22 9 9 25 5 22 22 17 7 18 31 15 22 22 17 7 18 31 15 22 22 17 7 18 31 15 22 22 17 7 18 31 15 22 22 17 7 18 31 15 22 22 17 7 18 31 15 22 22 22 17 7 18 31 15 22 22 22 17 7 18 31 15 22 22 22 17 7 18 31 18 18 18 19 25 22 22 17 17 18 18 18 18 18 18 18 18 19 19 9 9 9 15 16 6 6 6 30 11 7 7 22 22 17 7 18 18 18 18 18 18 18 18 18 18 18 18 19 18 18 18 18 18 18 18 18 18 18 18 18 18	12 7 13 221 21 21 20 25 25 25 25 17 16 31 15 18 8 34 17 12 9 9 9 9 10 16 6 6 22 22 12 10 23 23 25 25 25 25 17 16 20 10 25 25 25 25 25 25 25 25 25 25 25 25 25	3	13 10 13 22 33 2 9 27 24 17 18 31 20 13 35 18 12 9 9 14 16 7 30 12 2 33 2 9 27 23 36 20 13 37 2 9 27 24 17 18 31 31 2 2 33 2 9 27 24 17 30 2 9 27 24 31 3 2 9 27 24 31 3 2 9 27 24 31 3 2 9 9 27 24 31 3 2 9 9 27 24 31 3 2 9 9 27 24 31 3 2 9 9 27 24 31 3 2 9 9 27 24 3 3 3 2 9 9 27 24 3 3 3 2 9 9 27 24 3 3 3 2 9 9 27 24 3 3 3 2 9 9 27 24 3 3 3 2 9 9 27 2 3 3 3 2 9 9 27 2 3 3 3 2 9 9 27 2 3 3 3 2 9 9 27 3 3 3 15 2 9 9 27 3 3 3 15 2 3 3 3 15 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3		14 13 15 24 64 3 14 31 19 20 30 30 35 25 5 18 8 34 10 30 30 30 30 30 33 6 34 6 34 30 30 30 33 34 33 34 33 34 33 34 34 33 34 34 35 34 35 34 35 35 34 34 35 34 35 34 35 35 34 35 35 35 36 36 37 37 37 37 37 37 37 37 37 37 37 37 37

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co	verag			Time	Score	
		2	.8			×.		
	8	- 2	- 12	2	8	- 2	- 18	
	- 6	- 1	- 1	- 4	- 6	- 2	- 2	- 4
	- 8	8	8	8	8	8	8	8
agricola (20)		0	0	0	0.0	0.0	0.0	0.0
aimort can	- 34	24	- 34	- 34	2.5	18.5	26.1	26.1
barman (ba	- 14	- 0	14		0.3	0.0	2.0	18
Mocks (19)	28	20	28	28	2.2	19.5	29.1	29.1
childsmark con					0.0	0.0	0.0	0.0
data-network (20)	- 14	-nî	- 14	- 14	1.1	7.2	12.1	12.1
depot (27)	- 13	6	- 13	13	1.0	3.3	9.6	9.8
driverlog (20)	15	ž	15	15	1.1	4.3	10.5	10.7
elevators (70)	- 44	12	- 44	44	34	2.6	25.9	26.6
Boostile cam	- 72	10	- 76	6	0.3	0.0	0.8	0.8
freecell an	68	30	68	68	4.8	10.9	15.8	36.1
and can		- 73	19	19	11	4.2	101	12.2
and an	- 12	- 1	- 12	- 12	6.2	10		
Party Co.	- 2	- 2	- 2	- 2	0.0	- 12		
Buddhes (16)	- 6				10	- 22	10.0	10.7
rating (20)	- 22		10	10	1.0		10.9	10.7
sufference (stat)						11.0		
meconec (150)	111	100	100	100	114	41.7	121.0	42.6
Contract Contract	- 22	- 22	- 55	- 10	- 53	10.1	200	
inference (199	- 22	- 52	- 17	10	- 52	10.1		
mystery cus	- 22	12	- 12	12	12	12.0	100	10.0
nonrystery (30)	- 22	12	- 20	20	1.5		15.3	15.4
opensizers (100)	- 22	- 12	- 22	1	- 22	12.0	10.0	
organic (33)	- 2	- 2			0.5	6.0	- 51	0.1
organic-spin (20)				10		10		
purchance (50)	- 22	- 22	- 22		2.9	28.2	22.5	34.2
burgang (ac)	-12	- 1	- 13	-13	0.9	0.1	- M	54
pathways (50)		- 2			0.5	5.1		3.4
believe (act	- 12	42	48	48	3.7	22.8	35.6	33.4
petri-net (20)					0.0	0.0	0.0	0.0
bebos-ar (20)	- 25	14	- 25	25	1.8	10.4	19.0	18.7
pipes-1 (50)			18	18	1.3	5.1	12.1	12.0
bat-sumpg (20)	- 59	49	- 50	50	3.9	48.2	54.6	54.5
FOTOES (40)		- 7			0.6	6.6	8.1	8.1
satellite (36)	_ 2	6	- 7	- 2	0.5	5.5	7.3	7.2
scanalyzer (50)	38	- 7	- 33	35	2.7	5.7	19.6	21.2
snake (20)	12	6	12	12	0.9	2.5	7.6	7.4
sokoban (50)	50	33	50	50	3.8	19.6	39.5	39.9
spider (20)	15	.7	15	15	1.1	2.9	8.6	8.5
storage (36)	16	14	16	16	1.2	12.5	17.1	17.1
termes (29)	12	0	12	12	0.8	0.0	3.2	3.2
tetris (17)		3	10		0.8	1.3	5.4	5.5
tidybot (40)	25	18	- 24	25	1.8	5.8	15.3	15,4
tpp cm		7	8	8	0.6	8.0	8.9	8.9
transport (20)	- 34	20	.86	26	2.6	10.4	22.4	22,4
trucks (70)	-13	- 9	13	13	0.9	5.3	8.6	8.8
visital1 (m)	30	33	30	30	2.3	27.8	30.3	30.3
weedwerk (50)	-49	38	49	49	3.8	24.3	40.5	44.7
zenotravel (20)	-13	- 7	13	13	1.0	4.9	8.7	8.8
Sum (1827)	1156	766	1155	1159	86.8	539.8	900.4	908.7

					L(049)				
	Mind	14	5	-	ONS.	a	ø	7	τ
agricola	0	0	10	16	1	13	12		9
airport	23	- 35	- 36	- 33	1	21	13	- 14	17
borman	0	0	12	- 20	- 0	•	0	0	- 0
blocks	15	35	33	35	21	33	27	35	35
childenack	0	0	- 1	0	0	•	0	0	0
data-network	0		- 5	2	0		0	0	0
depet	- 4	14	18	18	- 6	- 5	- 5	1.5	13
driverlog	- 7	19	18	18	- 8	12	13	15	- 14
elevators	0	- 5	- 20	- 7	0		0	0	- 0
footile	0	0	2	1	0		0	0	0
freecell	30	-45	- 79	80	15	88	23	- 61	57
pod	0	20	20	20	0	28	0	0	0
grid	1	- 3	- 4	- 8	0	- 3.	0	- 3	- 4
ariterer .	8	28	- 20	20	8	24	- 10	28	- 20
biking	2	- 3	20	20	2	2	5	- 3	- 3
logistics	2	- 7	29	15	2	2	- 3	6	5
maintenance	0	14	11	14	0		0	0	0
miconic	35	159	150	150	21	159	145	159	150
movie	30	34	30	30	30	34	38	34	30
received	30	21	32	22	5	18	13	18	17
EPONO V	15	15	17	14		6	12	10	- 0
operata da	0	0		20	0	i i		2	- 6
pressie combrais			- 2						2
nonceptator.		- 12	- 20	- 18	- ñ		- 1	- 6	
neting	i ii		- 74		- ñ		- â		
rotheras	1		- 10	- i	1	1	- 4	- 4	1
page 1			- 20	- 20			- 14		
pigner .	1.11	- 55	- 11		1.22	1.5		- 55	
			- 50						
per la companya de la	1	- 51	- M	- 25	17	177	17	16	16
and he	1.2	- 12		- 15	1.1	1.12	17	17	
considered			- 12			1.1	- 6		- 6
scalary see			- 72		1.1		12		- 72
and the second s		- 2				1.4			
solution and a solution of the	1	12		10		1.1	14		- 2
dance		11				1.4	- 44		- 2
savege	1.2	12	12		14	12	- 7	12	
termes.			1		1.2			- 1	- 2
leas .		- 22	- 2	20	1		- 72	1	- 12
testing	2			- 22	2	14	2	2	2
Edybox.	1.3	12	10	- 20			- 2		- 3
66	2	- 19	- 23	29	2	15		- 12	19
Easycet				10					
Inacks	6	_ 2	15	. 2	6		- 2	- 2	- 2
visital	0	20	0	20	0	•	- 9	0	0
woodworking			2	- 4	1				
centravel	- 8	20	20	20	2	. *	- 9	13	- 14
SUM	399	775	933	965	345	651	558	631	632
$>h_K$					- 4	- 6	- 4	- 6	- 7
$>b_{\rm H}$					- 2	- 5	- 5	- 4	- 3
She					0	0			

Automatic Configurat	tion of Benchmar	k Sets for Cla	assical Planning
----------------------	------------------	----------------	------------------

	Hind Search		<u> </u>		A	1	a alMost						
Dominance Pri	ning Duplicate (hecking		Dom	nance Pro								36
tomain # $\exists_B \exists_B^{\prime\prime} \exists_B^{\prime\prime\prime} \exists_G^{\prime\prime}$	30 30 30 30 30 3	5 <u>6 56</u>	3a	ಸಕ್ಷ ಸ	z = a							0	ă.
uranet 20 9 9 5 5		1 1	15	14	4 4							Š.	- 3
riverbox 20 11 11 11 11	11 11 5 5	10 10	Lú.	13	15 15							15	-8
levators 30 6 6 9 1	16 16 0 0	10 10	10	ii –	22 13						1	-8	8,
loortile 40 2 2 2 1	2 2 0 0	0 0	10	10	10 10							1	- 54
reecell 42 0 0 0 0	0 0 0 0	2 2	1.1		2 2						a 18	÷.	8
ED 20 13 13 15 19	15 15 7 7	15 15	15	15	15 15	5 6	loverage						
			1.3				triverlog (20			1	15 15	13	14
femic 145 46 47 47 47	45 47 42 42	5 5	1.6	155	ns ns		levators-opt	08-strips	(30)		20 20	21	18
oMystery 20 20 20 20 20	19 20 16 16	16 16	20	20	20 26	i (devators-opt	11-strips	(20)		17 17	18	15
penSt14 20 1 2 1 1	2 1 2 2	2 2	2	2	1 1	i (loortile-opt1	1-strips	(20)		4 3		
SR 48 48 46 48	48 48 42 42	46 46	48	48	47 48	• •	reccell (NO)				55 65	- 11	- 17
overs 40 7 7 7 7	7 7 6 6	1 1	1.5	8	8 8		ed-opt14-so	ubs (20)			0 10	100	
Method 14 10 3 3 3 3 3	1 1 1 1	1 1	12	4	2 2		income (15)				18 27	28	27
ransport 28 10 11 13 14	15 15 0 0	ம் ம	12	12	<u>ніі</u> ій		omvskry-o	et I - stri	05 (20)		20 20	13	13
foodwork 26 7 7 7 7	7 7 7 7	7 7	16	16	16 11	1	arcprinter-0	8-strips	300	i	17 18	16	19
		5 7	12	12	12 12	1	sreprinter-o	pt11-stri	ps (20)	1	13 14	12	15
		67	87	87	87 87	1	ipesworld-n	ounkag	a (50)		22 22	22	23
	899 A	286	499	460	473 466	2	camlyzer-0	S-strips (30)		13 13	10	10
287787	777 8						canalyzer-o	911-510	ps (20)		10 10	7	
	888 7	7				- 1	nice optitis	support a	<u>"</u>		1 IA	- 11	
5.6.2.5.5.X	333 6 1	-				- 1	cara part ont	08-strine	con .		1 II	- 13	13
bellman - 7 5 5 4 5	5 4 4 1145 99	5				- 1	namport-opt	11-strips	(20)	. i	10 10	9	8
novelty 1 7 - 4 7 5 1	6 5 3 1153 112	í.				- 1	ransport-opt	14-strips	(20)		8 8	- 8	7
interval.1 7 8 - 4 5 4	5 4 2 1153 80	í —					voedworking	g-op108-	strips (3)	0) 3	20 20	19	26
		1.000	lane.		L C T D		voodworking	z-opt11-	strips (2)	0) [14 14	13	18
merva-10 7 8 4 - 6 4	4 2 3 1154 95	are	inv	mo	OVD	. 3	senotravel (2	0)		,	13 13	- 7	7
novelty-2 7 6 7 8 - 6	4 3 4 1157 105	28	27	23	29	- 1	ium (1827)			10	35 1042	974	1016
interval-100K 8 6 7 7 6 -	5 4 1 1156 113	28	- 27	28	28								
interval-1K 10 8 9 8 4 5	- 4 2 1157 110	12	12	14	13	- 1	1 12	12	12	12	10		- 63
interval-100 8 8 7 6 5 4	4 - 3 1157 106	Ч ń	- 14	13	l ú.	1	i 13	13	13	- 13	13		15
interval-10K 11 8 8 9 6 4	4 6 - 1159 112	5 22	22	20	22	2	2 22	22	22	22	22		24
	Freecell (80)	15	15	15	24	1	5 12	15	15	21	33		64
	Grid (5)	2	2	1	2		2 2	2	2	.2	2		3
	Hiking (20)	10	10	. 9	10		> 8	. 9	. 9	10	9		14
	Logistics (63)	27	- 27	25	27	- 2	5 25	25	25	25	27		34
	Mprime (35)	23	25	22	23	- <u>z</u>	3 23	22	22	- 22	24		31
	Nonestery (19)	16	16	14	17	- 6	14	18	19	16	19		20
	Operatoria (80)	31	- 31	- 31	31	- 3	30	31	31	31	31		34
	OrgSynth-split (20	15	15	14	15	1	4 10	15	15	15	15		10
	Parcprinter (30)	19	22	19	22	1) 19	22	22	18	20		30
	Parking (40)	9	9	6	9	10) 10	12	12	8	13		13
	Pegsol (36)	35	34	33	35	- 3	4 34	35	34	- 34	35		35
	Pipes-notank (50)	18	18	17	18	- <u>r</u>	7 17	18	18	17	18		25
	Pipes-tank (50)	12	12	10	12			12	12	12	12		18
	Proceeding and the Proceeding of the Proceeding	1 3								- 21			
	Satellite (b6)	14	12		6	1.1	í 11	16	16	- 16			- ÷
	Scanalyzer (30)	16	16	16	16	- 6	5 14	16	16	16	16	1	18
	Snake (20)	6	6	4	6		i 4	6	6	6	7		13
	Sokoban (30)	30	29	30	30	3	30	30	30	30	30		30
	Spider (20)	1 11	11	9	12	1	9	11	11	10	12		15
	Termes (20)	7	6	6	7		5 6	7	7	6	7	1	13
	Tidybot (40)	23	22	20	23	- 25	0 15	22	22	22	23	1	22
	VISIAI (40)	1 16	- 15	17	1 15	- 2	36	.56	36	14	36	1	30
	Woodworking (30)	19	22	19	19	- 2	22	20	20	- 22	20		30
	Amonavel (30)	1 13	13	12	1.13		- 12	15	45		12	1	
-1	Others (601)	331	331	331	331	- 33	331	331	331	351	331	1	.546
nina	Sum (1672)	1 856	865	821	873	86	850	884	883	858	914	1	1020

4/25

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Co	verage			Time	Score		1													
	_	×.	-8		_	8	4	_			_							8				_
	2	- 8	- 12	8	2	- po	1	8										Я.				
	- 8	- 2	- ē	1	- 6	- 2	- ē	10										a.				
		-												3		10	8	2	~			
agricola (20)					0.0	0.0	0.0	0.0						2		2	2	00			9	u
airport (50)	- 24	- 24	- 25	- M.	2.5	18.5	26.1	26.1			age	icela		0	0	00	16		13	12	11	9
Norks (10)		20	- 62		2.2	19.5	19.1	-			airg	NH.		23	- 35	- 36	- 33		21	13	- 14	17
childsnack (20)					0.0	0.0	0.0	0.0			bor			0	0	12	20	0	۰	0	0	0
data-network (20)	14	11	14	14	1.1	7.2	12.1	12.1			1 blo	eks.		15	35	35	35	21	33	27	35	35
depot (22)	13	6	13	13	1.0	3.3	9.6	9.8			chi	lázaci		0		- 1	0	- 0		0	0	0
driverlog (20)	15	- 7	15	15	1.1	4.3	10.5	10.7			dist.	- ectiv	ark.	0		. 5	2		•			0
elevators (70)	- 44	12	- 44	44	3.4	2.6	25.9	26.6			eep				12				1.3		13	13
Boortile (40)					0.5	0.0								1.4	- 72		12	- 2	14	14	12	17
rrecen (00)		- 20		10	- 22	10.9	22.6	39.1						1.4		- 77	- 1					
grid (3)	- 7	- 1	- 14	17	62	10	22	2.3			fm	sel.		20	-45	- 22	- 80	18		23	41	57
erioner (20)	- 6	6	8	8	0.6	4.8	7.0	6.8			800			0	29	20	20	0	28	0	0	0
hiking (20)	- 14	- 8	15	15	1.0	5.1	10.9	10.7			ph			1	- 3	- 4	. 8	0	3	0	- 3	- 4
logistics (67)	39	19	39	39	2.8	11.8	25.2	26.0			10	PCT		8	20	20	20	8	38	00	20	20
miconic (150)	144	133	144	144	11.2	80.6	131.6	131.9			14	ing –		2	- 3	20	20	2	- 7	- 5	- 3	- 3
movie (30)		20	- 20	20	- 2.4	42.7	42.4	2.0			big	inter		2	- 7	29	15	2	- 2	- 3	6	- 5
inprime (35)	- 27	24	19	19	2.5	12.6	17.8	29.7			FRA	index an	66	0	- 14		- 14	0	•	0	0	0
Doministery (Mr.	- 54	12	10	10	15	80	153	16.4			a new second	xek		55	150	150	150	21	170	145	150	150
openstacks cross	- 53	21	53	53	3.8	12.0	29.8	30.0			190	-		20	- 22	- 22	- 22	. 20		- 20	- 29	- 20
organic (30)	7	7	7	7	0.5	6.0	6.1	6.1			141			39	- 21		- 22	- 2		- 22		
organic-split (20)	10	6	10	10	0.7	1.9	4.2	4.2				any .		10	15	- 12				11		
parcprinter (50)	38	-34	38	38	2.9	28.2	34.1	34.2								- 2			1.4	- 6	- 6	- 6
parking (at)	- 13	1	13	13	0.9	0.1	54	5.4			100	coriste	,	6	- 12	- 20	18	- 0	1.	1		- ô
pathways (50)		- 2				2.1		24			100	h ing		0	0		0	0		0	0	0
pegad (30)		16	- 20		- 66	60	200	0.0			ped	ways.		- 4	5	- 80	. 8	- A	- 4	- 4	- 4	- A
Dines at the	25	14	25	25	1.8	10.4	19.0	18.7			199	part -		17	29	20	20	15	28	29	29	33
pipes-1 (5th	18	÷.					di		÷ .	B	D	9		12	22	23	27	12	13	15	20	17
per-small (50)	50	41				_		-		_		_		43	59	50	50	59	59	59	59	50
ECTEES (40)	8		-		_		****							6	21	26	- 25	- 6	14	36	16	16
satellite (36)	_ 2		arp	H (50)		2	5 25	- 24	- 24	- 24	- 24	- 24	- 24	<u>•</u>	15	n	12	- 7	15			. 9
scanatyzer (50)	- 22		data -	off one	nk (20)	- 3				- 22	- 22					- 22	- 20	2				
solution (20)		*	red	20.		- 7	ŝ	19	19	19	19	19	19		- 6			- 2	1.1			
souther con	- 18	٩.	hiki	100		- î	íű	12	12	- iii	- iii	12	12	1	10		19			14	1	- 6
storage con-	16	1	logis	tics98	(35)	- i	2 12	12	12	12	12	13	13	L ú	18	- 19	19	n	15		18	- 19
termes (20)	12	1	mice	nie (15	0)	14	4 143	144	144	144	144	143	144	6	10	14	15	5	4	- 2	2	
tetris (17)			mpri	ERE (35	•		30	30	30	31	30	30	30	0	29	9	20	2	. 11	19	2	18
tidybot (40)	25	1	myst	ery (38			5 15	19	19	19	19	19	19	5	5	8	14	5	12	5	5	5
tpp can	8	4	oper	stacks	08 (36)			- 24	- 11	- 11	- 12	- 23	- 23	- 3	19	- 16	20	0	1	- 7	8	2
tracks can	- 6	1	oper	and the	14		1 1		1	- 14	1	14	17	- 6	13	- 23	29	- 6	15	90	- 14	- 15
visital on	- 30	x	Parc	erinter	08 00	. 3	i 31	28	- 28	- 39	- 39	- 30	30	0	5	0	16	0	۰	0	0	0
weedwork (50)	- 49	3	parc	printer	11 (20)	2	20	18	18	20	20	20	20	6	_2	15	2	6	6	2	- 2	- 2
zenotravel (20)	13		park	ing-11	(20)		4 4	- 7	- 7	7	7	- 7	7	0	- 29	0	20	0	•	- 2		
Sum (1827)	1156	764	park	ing-14	(20)		4 4	. 6	. 6			.6	.6	1.			- 22	- 1	1.1	- 6	- 61	
Den (den)	11.00		pipe	world	not. (5	» 2	21	24	- 24	- 24	- 24	- 24	- 24					w.				an.
			pape	owona	-C (90)	- 1		- 10	10	10	- 10	14	10				-	- 4	6	-	- 6	7
			1010	aberner.	05 (32)	- 1		18	18	- 18	- 18	- 18	18					- 2	5	5	- i	à
			scan	alvzer-	11 00	- î	2 12	15	15	15	15	15	15					0	2	1	0	0
			snak	0 (20)			4 4	10	10	11	11	10	n.		-	-	-	-	-	-	-	-
			spid	tr (28)		1	2 12	- 14	14	15	15	15	15									
			term	es (20)			0 10	12	12	13	13	13	13									
			tetris	(17)				10	10	10	10	10										
			- udyt	NR-11	289	- 5	4 13 4 4	14	14	14	14	14	- 14									
			Internet	0010			6 10	- n	- 13	- 14	- 13	- 13	- ni									
			-					-				104										
			sam	(VZI) r domo	less come		537	596	590	609	635	613	610									
			-110	0.114	ana (900					100	100	100	100									
			- nonel	(1877)		107	1 1072	1131	1131	1144	1143	1140	1145									

Automatic Comparation of Denominary Sets for Classical Flamm	Automatic Confi	ouration of	Benchmark Se	ets for Classical	Plannin
--	-----------------	-------------	--------------	-------------------	---------

							В	lind	Searc	۸.	Deal					Dee		A		a riMan			1				
Domain		×.,	⇒ğ		, T	1	g	-1	-12	17	1 in	-1	1	2	14	-07	-117	-17									- 3
DataNet	20	9	-	-	1	-	9			9	5	5	3	3	14	14	12	14								2	+10
Depots	22		- 4		1.5		.4			4	2	2	4		1.7			- 3									Ĩ
Elevators	30	6			- 7		12	10		6	ő	ő	10	10	10	11	22	12							1	4	8.
Floortile	-40	2	- 1		- 4		2			2	0	0	0	0	10	10	10	10	•						÷.	8	- 14
Freecell	42	0					0			0	0	0	2	.2	1.1	.1	. 2	1	Ŀ.,					÷	12	1	2
Grid	- 5	11	- 11		- 11		"	- "		1	Lí.	- í - '	۲ <u>۳</u>	- 11	1 2	12	12	- 13		continge			_				
Logistics	63	24	2		2		26	- 25	- 3	26	22	22	24	24	34	34	- 35	- 34	1.5	friverlog	(20) cm/08_cm/c			15	15	13	14
Miconic	145	46	- 5		-5		47	-42		17	42	42 -	42	- 42	135	135	135	135		devators	opt11-strip	s (20)		17	17	18	15
OpenSt14	20	1	- 13		- 1		1	- 13		1	2	2	2	12	1 2	- 2	1	- 1	1	foortile-	pt11-strip	(20)		4	3	0	1
PSR	48	48	- 42		-46		45	-42		48	42	42	46	46	48	48	-47	- 42	<u>د</u> ا	freecell ()	40)			65	65	- 47	47
Rovers	- 40	- 7			- 2		7	- 3		21	6	6	1	- 1	8	8	8		: :	ged-opt1-	1-strips (20) 1505	•		15	110	109	113
Tidsbot14	10	- 5			- 3		á	- 5		51	5	÷.	ŝ.,	- 51	6	6	- 6	- 6		nprime (15)			28	27	28	27
Transport	28	10	- 1		-12		14	12	1	15	0	0	13	13	12	12	14	12		iomyster	y-opt11-st	ips (20)		20	20	13	13
Woodwork	26	_7			_3		2			21	1	1	3.		10	10	10	- 8		pareprint	ar-08-strips	(30)		17	18	16	19
ö							¥		~ ~				i-	67	87	87	87	8	1	Pipeswor	ki-notanka	20 (50)		22	22	22	23
Σ					2		8	×	8 8					286	459	460	473	-466		canalyze	r-08-strips	(30)		13	13	10	10
		- 5	k	1	1	k	1	2	19 19		- 8									icanalyze	r-opt11-st	ips (20)		10	10	.7	.7
			Į,	ł	8	¥.	ł	8	3 3		- 3	- 5							- 1	onite-opt	18-strips ()	30) 7)		8	10	- 13	
				-11	-8		-12		a a		_	_							- 1	ranaport	opt08-strip	s (30)		î4 -	14	- 13	13
bellmar	1		7	5	5	-4	5	5	4.4		1145	995								ransport-	opt11-strip	ts (20)		10	10	- 9	8
novelty	-1	- 7	-	4	7	5	3	6	5 3		1153	1125								ransport	opt14-strip	IS (20)		.8			- 2
interval	-1	- 7	8	-	4	5	4	5	4 2		1153	803	-						- 3	voodsor	king-opt08	-strips (2	m –	14	14	13	18
interval	-10	- 7	8	4	-	6	4	-4	2 3		1154	957	11	urb_	inv	m	1 6	ZD	. 1	centrave	1 (20)			13	13	7	7
novelty	-2	- 7	- 6	2	8	1	6	4	3.4		1157	1058		28	27	23		29	- 1	Sum (182	7)		1	135	1042	974	1016
interval	-100K	. 8		7	2		2	8	11		1156	1137		28	- 27	- 22		28	-								
interval	100		:	4	2	2	4	ā.	1.5		1157	1061	1	7	7	- 1		7		ĩ -	i 'i	7	- 7		10		13
internal	108			6	÷.	ŝ	2	2	2.1		1109	1125		13	14	12		14	1	3 1	13	13	13		13		15
	TUR												1	22	22	- 20		22	- 2	2 2	2 22	22	22		22		24
									Grid	(5)	(10)			2	2			2		2 3	2 2	2	- 12		2		3
									likir	18	(20)			10	10	ŝ		10		9 i	3 9	9	10		9		14
									Logie	tic	3 (63)			27	27	- 25		27	2	5 2	5 25	25	25		27		34
									Mpri	me	(35)			23	25	- 1		17	- T	3 2	22	17	25		24		31
									Nom	¥50	ery (2	05		16	16	14		17	- it	5 14	18	18	16		18		20
									Open	eta	cks (I	KO)		31	31	- 31		31	3	1 31) 31	31	- 31		31		34
									OrgS	2m	th-spl	it (20)		15	15	14		15	- 1	4 10) 15	15	15		15		10
									Parep	enn Dat	ther (3	0)		19	22	- 19		22	- 1	8 1	22	22	18		20		30
									Peese	яĉ	36)			35	- 34	33		35	3	4 3	35	34	34		35		35
									Pipes	-00	stank	(50)		18	18	13		18	1	7 1	7 18	18	17		18		25
									Pipes	-12	nk (51	7)		12	12	- 19		12		1	2 12	12	12		12		18
									Post		gnme	m (20)	١.										- 2				0
									Satell	lite	(36)			8	12	- 6		8	1	ά <u>μ</u>	15	15	10		14		8
									5cana	aly.	zer (3	0)		16	16	- 16		16	13	5 14	16	16	16		16		18
									šnak	0 (î	50)			<u>.</u>	6	- 2			- 2	4	6	6					13
									Soide	oar r c	1(30) 205		Т	8	29	- 34		12	- 1	0 34	30	30	39		30	1	30
									Ferm	esi	(20)		Т	7	6	- ê		7	- 1	6 4	5 7		6			1	13
									Fidyt	x	(40)		Т	23	22	20		23	25	0 1	5 22	22	22		23	1	22
								1.	Visit	AII	(40)		Т	16	- 15	- 13		15	- 3	6 34	36	36	14	1	36	1	30
									mooc Zeno	y sa q trar	oriung vel (2	r (30) N	Т	13	- 13	13		13	- 25	9 21 2 13	: 20 13	13	13	1	20	1	13
								П	Other	rs (601)		Ť3	31	331	331	i.	331	33	1 33	331	331	331	÷	331	i -	346
nninc	1								Sum	(14	(72)		18	56	865	821	i.	873	86	9 85) 884	883	858	i	914	i	1020

4/25

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Empirical Evaluation – Examples from HSDIP'20

		Ce	werage		Time Score																								
		4	-2		3 4					8																			
		1 2	3 8		2 2	8				8							Bli	nd Search		_		A	- 10 I D	Los .		-			
		ųį	- Ř 19	- U	Ř Ř	12				8					Do	minance I	Presi	g Duplicate C	hecking		Domir	ance Pru							34
									3	1 X			Domai		≾a ≾%	31 ¹⁷ 3	97 -	St St In Sh :	ೊ ≾ೇ	30.3	dT ⇒t	T ≾7						-	1
	agricola (20)		0 0	0.0	0.0 0.0	0.0			2 6 2	a os u u	a u		DataN	4 20	9 9	5	9	9 9 5 5	3 3	14	14	12 14						1	24
	airport (50)	34 24	M M	2.5	18.5 26.1	26.1		arricela	0 0 10	16 1 12 1	2 11 9		Depote	22	3 3	4	4	4 4 2 2	4 4	7	7	7 7						5	÷.
	barman (34)	4 0		0.3	0.0 2.0	1.8		sirport.	23 35 36	33 1 21 1	3 14 17		Driver	og 20	н н	11	11	11 11 9 9	10 10	13	13	13 13						8	CI.
	blocks (39)	28 20	18 18	2.2	19.5 29.1	29.1		borman	0 0 12	20 0 0	0 0 0		Elevate	ni 30	6 6		12	16 16 0 0	10 10	10		22 13					7	- 12	
	childsnack (20)			0.0	0.0 0.0	0.0		blocks	15 35 35	35 21 33 2	2 35 35		Floorti	e 40	2 2	2	2	2 2 0 0	0 0	10	10	10 10						3	1
	data-network (30)	14 11	14 14		7.2 12.1	12.1		childrende	0 0 1	0 0 0	0 0 0		Freece	1 42	0 0	0	0	0 0 0 0	2 2			2 2	-				a 8	10	8
	depot (22)	13 6	13 13	1.0	3.3 9.6	9.8		data activity	0 1 5	2 0 0			GED	20	13 13	15	15	15 15 7 7	18 19	18	15	15 15	Cove	calles					
	deverlog (20)	10 .7	15 15	1.1	4.3 10.5	10.7		Acres		18 6 3			Ond							1.4			drive	dog (20)			15 15	13	14
	CREVINCES (50)	** 12		- 22	2.6 25.9	26.0		delegation	2 10 10		3 15 14		Logos	0 00			-	2 20 24 24	8 8	1.2	а.	8	eleva	ters-opt08-stri	os (30)		20 20	21	18
	Boortine (40)	. 0		0.3	0.0 0.8	0.8		and the second					MKOB	g 145	40 47		**	45 47 42 42	<u>12</u> 12	1.00	135 1	35 135	eleva	ters-ont L1-stri	os (20)		17 17	18	15
	freecell (80)	68 30	68 68	4.8	10.9 33.8	36.1		elevation	0 5 20				Down	tery 20	20 20	20	20	19 20 16 16	10 11	1.0	20	20 20	floort	ile-cot11-strip	6 (20)		4 3	0	1
	Berg (3a)		10 10	1.3	4.7 12.1	12.2		factoria de la constante de la					DSD	44			÷.,		ŵ ŵ	1.4	÷.	ė a	freeo	all (ND)			65 65	47	47
	Barra (2)			0.2	10 22	2.3		ineces					Report				- N	7 7 6 6		1.7			and a	ot14-strips (2)			15 15	19	19
	Budden (se)			0.6	4.8 7.0	0.8		fee	0 20 20	20 0 20	0 0 0		Satella	· 36	1 1				1.1	1.2	-	9 7	mico	aic (150)		1	10 110	109	113
	mang (20)		10 10	1.0	5.1 10.9	26.0		free	1 2 2 2				Tidsbe	14 10	3 3	3	3	3 3 3 3	3 3	6	6	6 6	mprie	ae (35)			28 27	28	27
	software (ex)				11.0 20.2			Tables.	5 20 20	20 0 20 1			Transe	H1 28	10 11	13	14	15 15 0 0	13 12	12	12	14 13	nomy	stery-opt11-st	rips (20)	a 1	20 20	13	13
	meetine (150)	30 30	10 10	14	41.7 41.4	131.3		hiking	2 3 20	20 2 7	5 3 3		Woody	ork 26	7 7	2	7	7 7 7 7	7 1	16	16	16 17	PART	rinter-08-strip	a (30)		17 18	16	19
	merime (15)	29 24	19 19	- 23	19.1 26.5	26.7		politice.	2 7 29	15 2 2	3 6 5		Z4					n ni / /	1 1	12	12	12 13	parcp	rinter-opt11-s	trips (20)	0	13 14	12	15
	repeated (25)	10.11	19 19	14	12.6 17.8	17.0		maintenance	0 14 11	14 0 0	0 0 0		0				₩.	M	6	87	87	87 87	pipes	world-notanka	420 (50)		22 22	22	23
	Domination (30)	36 12	10 10	15	80 153	15.4		miconic	35 159 159 1	50 71 150 14	6 150 150		Σ			- 2	N N	488	284	499 -	460 - 4	73 466	ACT NO.	lyzer-08-strip	4 (30)		13 13	10	10
	conceptantia (and	43. 31	63 63	11	120 28.8	33.0		ERCY 10	30 30 30	39 39 39 3	0 30 30		-		1 6 1	4 H X	- 64		_				50000	lyzer-cot11-st	rips (20)		10 10	7	7
	opennic con	3.3		0.5	60 61	61		mprime	20 21 32	22 5 18 1	3 18 17				15	225	2.1	11 E E E E	8				make	-opt18-strips (201	· · · ·	13 12	13	13
	comparies and in case	10 6	10 10	0.7	10 41	1.1		mystery	15 15 17	14 9 6 1	2 10 9				705	335	111						tetris	-cot14-strips (17)		11 10	11	11
	precedence (20)	38 14	38 38	2.0	28.2 14.1	34.2		openstacks	0 0 2	20 0 1	8 2 5				.0 6		- 44	144 6 6					transp	ort-opt08-stri	es (30)		14 14	13	13
	mething cars	10 1	13 13	0.9	0.1 5.4	5.4		organic-symboli	3 3 2	3 3 3	3 3 2		bell	man	- 7	5 5 4	5.5	5 4 4 1145 99	5				transi	cort-opt11-stri	es (20)		10 10	9	8
	outher reaction	A A	5 5	0.1	51 66	5.4		porcprinter	0 12 20	18 0 0	1 0 0			day 1		4 7 4							transi	cont-opt14-stri	05 (20)		8 8	8	7
	mented ctop	48 42	48 48	3.7	22.8 35.6	35.4		parking	0 0 7	0 0 0	0 0 0		101	any-1	1.11	111		5 5 5 1155 112					WOOK	working-cet0	A-strips ((30)	20 20	19	26
	petri-net (20)		0 0	0.0	0.0 0.0	0.0		pathways	4 5 10	8 4 4	4 4 4		inte	real+1	7 8	- 4 3		5 4 2 1153 80					wood	working-opt1	1-strips d	(20)	14 14	13	18
	DEDGE-BH (50)	25 14	25 25	1.8	10.4 19.0	18.7		pegiad	17 29 29	diamont in	a					4 - 6	4.4	4 2 3 1154 95	7 arb	inv	rnd	GZD	zenot	ravel (20)			13 13	7	7
	pipes 4 (50)	18 1			di	55	E I	01	12 22 23	Collection	Comp	PDBs 2	scorp LN	out SY	MBA"	78 -	6 4	4 3 4 1157 105	1 28	27	23	20	Parent .			10		0014	1014
	per-small (So)	50 4							47 50 50	s bipartite	105	73	182	180	59	776	1.1	5 4 1 1156 113 ¹	28	27	28	28	50m	(1947)		10	35 1042	974	1016
	FOTOTES (40)	8			+11	•	+110	+10 +10	6 21 26	bd-bipartite	96	78	137	107	72			4 3 1153 110	12	12	12	11	12	12 12	- 17	- 121	<u> </u>		- 14
	satellite (Mo	7 1	airport (%		25 2	5 24	24 24	24 24 24	6 15 27	chain 0.1	159	121	105	372	292			- 4 2 H57 H0	2 2	17	- 61	1	12	7 7	- 12	i 191		ň I	- 63
	scanalyzer (90)	35	data-netv	eek (20)	13 1	3 14	14 14	14 14 14	4 20 18	Channel of the					202	765	4.4	4 – 3 1157 106	Ч ú	- 14	- ú l	- 14	- 12			(ú.)			16
	snake (20)	12 (freecell ()	105	42 4	3 60	60 61	61 59 60	3 4 5	chain 0.5	349	321	389	339	288	896	4.4	4 6 - 1159 112	5 22	22	20	22	22	22 22	- 22	1 2			24
	sokoban (50)	50 X	ged (20)		15 1	5 19	19 19	19 19 19	6 13 19	chain 0.25	409	345	444	428	309			Encodell (80)	1 15	15	15	24	16	12 15	15	6 51	1 7	i	64
	spider (20)	15	hiking (3)	0		1 12	12 13	13 12 12	1 12 9	chain 0.75	320	304	391	333	261			Child (D)	1 1	14	- 11								
	storage (36)	16 1	logistics5	8 (35)	12 1	2 12	12 12	12 13 13	14 18 19	complete	123	117	153	129	110			Hiking (20)	l liô	- iô	- 61	10	ő	÷ ő	. ő	i nël	1 1	ő I	
	lermes (29)	12 (miconic (150)	144 14	3 144	144 144	144 143 144	0 10 14	day 0.1	113	0.4	170	166	22			Logistics (67)	1 27	27	- 26	22	25	25 25		2 25		<u>.</u>	- 12
	tetris (17)		mprime (55)	- 30 - 3	0 30	30 31	30 30 30	0 20 9	ung o.r				100				Maxima (10)	1 22	25	22	22	22	21 22		1 51	1 5		
	tidybot (40)	25 11	mystery	38)	18 1	8 19	19 19	19 19 19	5 5 8	dig 0.5	63	43	108	121	25			Mastery (19)	16	17	15	17	17	17 17	17	i 11	1 1		19
	tpp cm		openstact	3-08 (30)	20 2	22	24 24	22 25 25	3 19 16	2 dag 0.25	66	50	134	108	15			Normal Control of Control	1.16	16		14		14 14					20
	transport (20)	34 2	openstact	(8-11 (28)	15 1	5 17	10 10	17 18 18	6 13 23	dag 0.75	73	62	113	110	32			Domestocky (20)	1 21	21	- 57	- 11	11	30 31	10	2 - 10 I		2	- 14
	trucks (30)	12.2	openstact	3-14 (20)		2.2		4 4 4	0 5 0	dichain	382	341	416	391	296			Opensiaties (80)	1.6	10	- 61	10	- 62	10 15	- 64	i 61	1 1		10
	visitali (at)		parcprint	27-08 (30)		20	18 39	30 30 30	6 9 15	finh	160	331	471	305	308			Bergayintespin (20)	10	- 22	10	- 22	10	10 22	- 22	1 10		š.	30
	WOODWORK (50)	22.0	parties.	Low			1 1	1 1 1	0 20 0	IOIK		241		393	308			Parking (40)	1 13		16		10	10 12	12	i 191	1 1	. I	11
	2220072021 (20)	13	parting-	1 (20)	- 2.3	: :	1 1	1 1 1	1 1 2	inverted forb	436	393	356	386	357			Parking (40)	1.4		- 12		10	10 12		:			10
	Sum (1827) 1	156 76	parting	Manage and				N N N	8 20 20	polytree 0.1	325	253	386	343	224			Pegsot (36)	1 20	34	- 22	30	17	34 20		: <u>31</u>	1 7		35
			ninerror	Mar units	° 11 1	1 11	11 16	16 14 16	399 775 933 1	polytree 0.5	323	269	373	338	254			Pipes-nounk (50)	1 13	13	- 16	10	- 11	1/ 10	10	6 161		8	10
						1 10	10 9	9 9 9		miliano 0.2	167	269	405	270	245			Pipes-tank (20)	. 1 12	10	10	10	12		. 12	: "1			10
			acanatica		13 1	5 18	18 18	18 18 18		polyate 0.2		200		400	200			Processing internation	1 2		- 61					(<u>7</u> 1			
			scanalyze	s.11.00	12 1	2 15	15 15	15 15 15		posyuree 0.7:	303	270	421	408	200			Rovers (40)	1.2		- 21	- 2							~
			anaka can		4 1	4 10	10 11	11 10 11		random 0.1	54	38	167	95	32			Salenne (56)	1.3	14			12	13 15	15	10			
			seider co		12 1	2 14	14 15	15 15 15		random 0.5	57	52	117	88	34			Scanalyzer (30)	16	16	16	16	15	14 16	16	10		2	18
			termes (2		10 1	0 12	12 13	13 13 13		random 0.25	97	89	153	129	81			Shake (20)	1.2							: <u>.</u>		41	10
			petris (17)		8	\$ 10	10 10	10 10 10		1 0.20					25			Sokoban (30)		29	- 30	30		30 30					30
			tidybet-1	1 (20)	12 1	3 14	14 14	14 14 14		ranoom 0.75								Space (20)	1 1		21	12	- 5	2 11	- 5	10	1 1	<u> </u>	15
			tidybet-1	4 (20)	4 3	4 8	8 9	9 9 9		star 0.1	226	130	272	224	125			Termes (20)	1.2	6	6	.7	6	6 7		6		21	13
			transport	(21)	10 1	0 13	13 13	13 13 13		star 0.5	257	144	333	272	139			Tidybot (40)	23	22	20	23	20	15 22	22	22	1 2	2	22
					\$36 \$7	1 606	\$06 600	602 605 610		star 0.25	219	114	290	248	115			VISIAII (40)	16	15	- 17	15	- 20	36 36	36	14	1 2	2	30
			other dee	anine com	S16 61	5 515	515 515	515 515 510		star 0.75	172	132	263	199	134			woodworking (30)	1 19	22	- 12	19	20	22 20	20	22	2		
			0000000	anno (900)		0.000				trees.	364	277	411	402	245			Zenotravel (30)	13	13	- 12	13	12	12 13	13	2 13		4	13
			total (1823	9	1071 107	2 1131 1	131 1144	1143 1140 1145			304							Others (601)	331	331	331	331	331	331 331	331	331	33	1	346
Automa	tic Con	fiai	irati	on	of B	on	chm	ark S	ote for	1226	1831	De	ímii	100 A	4416			Sum (1677)	1 966	965	971	972	860	250 224	993	0.00	01	41	1020
Autollia	00 000	ngu	arau			201	CIIII	an N J	CI3 101	CidSt	sical	- 10		iy –				50m (1672)	1 800	\$65	021	673	900)	850 884	883	- 638 J	, 91-	• 1	1020

Evaluation

Conclusion

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

Evaluation

Conclusion

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

Having a standard evaluation setting is generally beneficial:

- Reproducibility
- Interpretability
- Avoids hand picking results

Evaluation

Conclusion

Empirical Evaluation – The ICAPS/IPC Way

The ICAPS/IPC Way

- Measure coverage
- Time limit 30 minutes
- Memory limit 2-8 GB
- Use the benchmarks from the International Planning Competition

Having a standard evaluation setting is generally beneficial:

- Reproducibility
- Interpretability
- Avoids hand picking results

The ICAPS Way	Benchmark Design Principles ●○○○○○	Benchmark Configuration	Evaluation 00000	C
Outline				

- 2 Benchmark Design Principles
- Benchmark Configuration

4 Evaluation



Benchmark Design Principles

Benchmark Configuration

Evaluation 00000

Conclusion

The diversity in the IPC Benchmark Set



Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

So, What's Wrong with the IPC Benchmark Set?

		IPC	
	L	D	0
Nomystery (20) Rovers (40) Woodworking (50)	11 40 50	20 40 50	12 40 50
Total	101	110	102

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

So, What's Wrong with the IPC Benchmark Set?

		IPC	
	L	D	0
Nomystery (20) Rovers (40) Woodworking (50)	11 40 50	20 40 50	12 40 50
Total	101	110	102

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

- Different number of instances per domain
- Instance scaling: too easy, too hard, and not smooth

Benchmark Configuration

Evaluation

Conclusion

So, What's Wrong with the IPC Benchmark Set?

	IPC				Ν	ew'14	ł
	L	D	0	_	L	D	0
Nomystery (20) Rovers (40) Woodworking (50)	11 40 50	20 40 50	12 40 50		25 22 18	30 18 27	24 21 30
Total	101	110	102		65	75	75

Table: Coverage of LAMA (L), Decstar (D) and OLCFF (O)

- Different number of instances per domain
- Instance scaling: too easy, too hard, and not smooth

 \rightarrow Experiments on some domains of the IPC benchmark set may not observe any difference between planners even if it exists!

The ICAPS Way Benchmark Design Principles Evaluation 000000

Non-Smooth Scaling



The ICAPS Way	Benchmark Design Principles ○○○○●○	Benchmark Configuration	Evaluation 00000	Conclusion
Smooth Sca	aling			



The ICAPS Way	Benchmark Design Principles ○○○○○●	Benchmark Configuration	Evaluation	Conclusion
Contribution	า			

The ICAPS Way	Benchmark Design Principles ○○○○○●	Benchmark Configuration	Evaluation	Conclusion
Contributio	า			

Smooth scaling from easy to hard instances:

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration	Evaluation 00000	Conclusion
Contributio	n			

- Smooth scaling from easy to hard instances:
 - Easy: solvable by any planner that anyone would compare against (baseline)
 - Hard: out of reach of current existing planners within a reasonable time limit

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration	Evaluation 00000	Conclusion
Contributio	n			

- Smooth scaling from easy to hard instances:
 - Easy: solvable by any planner that anyone would compare against (baseline)
 - Hard: out of reach of current existing planners within a reasonable time limit

2 Minimize bias towards/against planners used

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration ●○○○○○○	Evaluation 00000	Co 00
Outline				

2 Benchmark Design Principles

3 Benchmark Configuration

4 Evaluation



Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Example Domain: Barman



Instance Generator

Benchmark Configuration

Evaluation

Conclusion

Instance Generation Problem

Input:

- o domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

Benchmark Configuration

Evaluation

Conclusion

Instance Generation Problem

Input:

- o domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

 $\text{Generate instances} \rightarrow \text{Compute/Estimate runtimes} \rightarrow \text{Select instances}$

Benchmark Configuration

Evaluation

Conclusion

Instance Generation Problem

Input:

- o domain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

 $\text{Generate instances} \rightarrow \text{Compute/Estimate runtimes} \rightarrow \text{Select instances}$

How to avoid bias wrt. the set of considered planners?

Benchmark Configuration

Evaluation

Conclusion

Instance Generation Problem

Input:

- odomain
- instance generator
- a baseline planner
- a set of state-of-the-art planners

Output: set of instances with a good scaling

Generate instances \rightarrow Compute/Estimate runtimes \rightarrow Select instances

How to avoid bias wrt. the set of considered planners?

Output: set of linear scaling of parameters for the generator that produce a good scaling in runtime

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration	Evaluation	Conclusior
0000		0000000		

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes	cocktails	shots	ingredients
Linear Attributes	$b \in [1,6]$	$b \in [1, 5]$	$v \in \{3, 4, 5\}$
	$m \in [1, 5]$	$m \in [0, 5]$	
		+ cocktails	

Enumerated attributes

Гhe	ICAPS	Way
200		

Benchmark Configuration

Evaluation

Conclusion

Sequences of instances

User specifies characteristics of the generator parameters:

	cocktails	shots	ingredients
Linear Attributes:	$b \in [1, 6]$	$b \in [1, 5]$	$v \in \{3, 4, 5\}$
Numeric value	$m \in [1, 5]$	$m \in [0, 5]$	
Increase size of the task		+ cocktails	

• User specifies ranges for the base value (b) and slope (m)

Enumerated attributes

The	ICAPS	Way

Benchmark Configuration

Evaluation

Conclusion

Sequences of instances

User specifies characteristics of the generator parameters:

	cocktails	shots	ingredients
Linear Attributes:	$b \in [1, 6]$	$b \in [1, 5]$	$v \in \{3, 4, 5\}$
 Numeric value 	$m \in [1, 5]$	$m \in [0, 5]$	
Increase size of the task		+ cocktails	

Enumerated attributes:

 User specifies ranges for the base value (b) and slope (m)

- Finite set of values
- Fixed in the sequence

Benchmark Configuration

Evaluation

Conclusion

Sequences of instances

User specifies characteristics of the generator parameters:

Linear Attributes:

- Numeric value
- Increase size of the task
- User specifies ranges for the base value (b) and slope (m)

Enumerated attributes:

- Finite set of values
- Fixed in the sequence

cocktails	shots	ingredients
$b \in [1, 6]$ $m \in [1, 5]$	$b \in [1, 5]$ $m \in [0, 5]$ + cocktails	$v \in \{3, 4, 5\}$

Our system may select sequences like:

(b = 5,	(b=1,m=0,	(v = 3)
m = 1.34)	+cocktails)	
5	6	3
6	7	3
7	8	3
9	10	3
10	11	3
11	12	3
13	14	3

The	ICAPS	Way

Benchmark Configuration

Evaluation

Conclusion

Optimization Process

- Generate candidate sequences that scale smoothly
- Choose selected (sub-)sequences to include easy to hard instances

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration ○○○○○●○	Evaluation 00000	Conclusion
Sequence	Optimization			

We use SMAC to optimize the value of b, m and v for each parameter

- Measure instance difficulty as the best runtime by any planner (time limit 180 seconds)
- Penalty based on smoothness of scaling difficulty (ideally by a factor of 1.5 to 2)

Runtimes:10.36, 15.41, 18.9, 28.02, 29.27, 68.01Ratios:1.48, 1.22, 1.48, 1.04, 2.32Penalties:0.02, 0.54, 0.02, 0.91, 0.13Total penalty1.62

Sequence Selection

MIP encoding to select sequences satisfying hard constraints:

- There are 30 instances
- (Easy) Baseline solves at least one instance in less than 30 seconds
- (Hard) Sub-sequences go from easy ($\leq 180s$) to hard (> 2000s)
- (Diverse) Don't repeat the same parameters more than twice

and soft constraints:

- (Easy) Baseline solves 2 to 6 instances under 30 seconds
- (Easy) State-art planners solve 8 to 15 instances under 180 seconds
- (Hard) All sequences end in a very hard instance
- (Diverse) Don't repeat the same parameters more than once
- (Smooth) Minimize penalty of selected sequences

The	ICAPS	Way

Benchmark Configuration

Evaluation

Conclusion

Outline

The ICAPS Way

- 2 Benchmark Design Principles
- Benchmark Configuration





Experiments

Compare our new benchmark sets against the IPC

- 26 domains
- Satisficing and Optimal track
- 2 new benchmark sets that differ on the "training set":
 - New'14: using planners up to 2014
 - New'20: using all available planners
- Evaluation based on planners from IPC'18

The	ICAPS	Way

Benchmark Configuration

Evaluation

Conclusion

Evaluation Criteria

How to evaluate the quality of a benchmark set?

How to evaluate the quality of a benchmark set?

- Coverage range: generally better if all planners solve some instance and no planner solves all instances
- Comparisons: number of pairs (X, Y) of planners, such that coverage(X) ≠ coverage(Y)

How to evaluate the quality of a benchmark set?

- Coverage range: generally better if all planners solve some instance and no planner solves all instances
- Comparisons: number of pairs (X, Y) of planners, such that *coverage*(X) ≠ *coverage*(Y)

Goodhart's law: "When a measure becomes a target, it ceases to be a good measure." – Marilyn Strathern

 \rightarrow Comparisons is a useful metric to compare benchmarks but not a metric to optimize for (would introduce bias towards the set of planners)

Benchmark Design Principles

Benchmark Configuration

Evaluation

Conclusion

Results

		cove	erage ra	nge	com	paris	ons			cove	rage ran	ge	com	paris	ons
Optimal	#IPC	IPC	'14	'20	IPC	'14	'20	Satisficing	#IPC	IPC	'14	, 20	IPC	'14	'20
barman	34	4-11	9-13	9-12	12	21	19	barman	40	39-40	7-25	9-30	7	24	27
blocksworld	35	18 - 30	5-12	5-12	18	24	24	blocksworld	35	35-35	7-24	4-22	0	27	28
childsnack	20	0-6	9-20	6-21	12	18	22	childsnack	20	1 - 20	14-30	2 - 19	27	25	28
data-network	20	6-14	5-12	5-16	27	25	27	data-network	20	9-19	10-30	13-30	24	27	25
depot	22	5-14	9-25	8-16	26	26	24	depot	22	21-22	12 - 20	11 - 26	7	27	22
driverlog	20	7-15	6-30	5 - 18	22	26	25	driverlog	20	20-20	29-30	9-19	0	12	24
elevators	50	28-44	7-14	10 - 18	26	26	23	elevators	50	49-50	30-30	30-30	7	0	0
floortile	40	16-34	9-18	8-17	21	21	22	floortile	40	4-40	1 - 12	1 - 11	17	25	24
grid	5	1-3	6-26	4-21	19	28	27	grid	5	5-5	4-20	9-21	0	26	24
gripper	20	8-20	11-30	11 - 30	7	7	7	gripper	20	20-20	26-30	26 - 30	0	7	7
hiking	20	12 - 18	7–9	5-16	23	15	25	hiking	20	10-20	2-22	3-26	24	28	27
logistics	63	13-34	5-17	5 - 14	27	27	25	logistics	63	51-63	5-30	5-26	17	27	26
miconic	150	56-142	4-28	3-30	25	27	28	miconic	150	150-150	30-30	30-30	0	0	0
nomystery	20	8-20	3-27	5 - 21	18	28	27	nomystery	20	12-20	19-30	2 - 30	23	18	26
openstacks	130	42-71	4-11	3-7	24	18	7	openstacks	160	99–160	12 - 21	14-23	21	27	25
parking	40	0-15	11 - 18	12 - 21	28	24	23	parking	40	36-40	14-20	13-16	7	24	21
rovers	40	6-13	4-26	6-19	25	22	7	rovers	40	38-40	10 - 22	6-30	7	26	27
satellite	36	7-14	8-30	4-27	22	25	26	satellite	36	26-36	5-30	6-14	23	17	23
scanalyzer	50	21-33	6-16	7-15	27	24	24	scanalyzer	50	48-50	9-16	13 - 14	12	21	12
snake	20	7-14	5-20	7-19	22	24	21	snake	20	3-17	6-30	5 - 14	27	28	26
storage	30	15 - 18	9-25	2 - 19	21	27	26	storage	30	21-30	6-26	7-17	26	27	26
tpp	30	7-20	7-30	2-7	24	24	21	tpp	30	29-30	10-26	6-21	15	27	27
transport	70	24-35	5-30	8-19	21	18	22	transport	70	65-70	22-30	15-23	7	24	26
visitall	40	12 - 30	6-21	5-20	27	27	27	visitall	40	36-40	4-30	4–29	7	24	26
woodworking	50	38-50	16-25	10-14	22	26	24	woodworking	50	28-50	6-30	5-30	13	27	27
zenotravel	20	8-13	6-30	3-13	23	26	28	zenotravel	20	20-20	6-29	5-17	0	23	25

Benchmark Configuration

Evaluation

Conclusion

Highlight: SAT track

	comparisons				
Satisficing	IPC	'14	'20		
gripper	0	7	7		
miconic	0	0	0		
elevators	7	0	0		
blocksworld	0	27	28		
driverlog	0	12	24		
grid	0	26	24		
zenotravel	0	23	25		
barman	7	24	27		
depot	7	27	22		
parking	7	24	21		
rovers	7	26	27		
transport	7	24	26		
visitall	7	24	26		

The I	CAPS	Way

Benchmark Configuration

Evaluation

Conclusion

Outline

The ICAPS Way

- 2 Benchmark Design Principles
- Benchmark Configuration

4 Evaluation



The ICAPS Way	Benchmark Design Principles	Benchmark Configuration	Evaluation	Conclusion ○●
Conclusion				

- New tool to automatically select instances
 - Our tool consistently generates well-scaled instance sets that are useful to evaluate current planners
- New benchmark set significantly better than the IPC benchmark set, specially in the SAT/AGL track

The ICAPS Way	Benchmark Design Principles	Benchmark Configuration	Evaluation	Conclusion ○●
Conclusion				

- New tool to automatically select instances
 - Our tool consistently generates well-scaled instance sets that are useful to evaluate current planners
- New benchmark set significantly better than the IPC benchmark set, specially in the SAT/AGL track
- We need your feedback!
 - Do you find the results of our tool useful?
 - Is there any reason to prefer the IPC set over our new one?
 - Are there any constraints that we should take into account (in general or for specific domains)?