Genetic Programming Benchmarks Survey 2012

1. In your work on Genetic Programming, do you run experiments to measure algorithm performance?

	Response Percent	Response Count
Yes, in every paper.	**50.0%**	**39**
Yes, in almost every paper.	30.8%	24
Yes, sometimes.	15.4%	12
No, I only do theoretical work, or I only use experiments for purposes other than measuring performance.	3.8%	3
answered question		**78**
skipped question		**1**

2. What types of problems do you use? Please check all that apply.

	Response Percent	Response Count
Well-known Genetic Programming benchmark problems.	**69.6%**	**55**
Problems from the Genetic Programming literature which are rarely used as benchmarks.	21.5%	17
Problems developed by me or my group.	53.2%	42
Problems taken from other fields, not commonly used in GP.	39.2%	31
Real-world problems.	65.8%	52
Other (please specify)	8.9%	7
answered question		**79**
skipped question		**0**

3. When running experiments to measure performance of a new technique, do you use existing techniques as controls?

	Response Percent	Response Count
Yes, I run experiments using existing techniques.	**63.3%**	**50**
Yes, I rely on existing reported results using existing techniques.	19.0%	15
No, I don't use control experiments.	7.6%	6
Not applicable.	10.1%	8
answered question		**79**
skipped question		**0**

4. If you run control experiments using existing techniques, do you re-implement existing techniques, or use code made available by others for this purpose?

	Response Percent	Response Count
Re-implement existing techniques.	32.5%	25
Use third-party code for existing techniques.	**44.2%**	**34**
Not applicable.	23.4%	18
answered question		**77**
skipped question		**2**

5. Does the GP community's current experimental regime have any disadvantages?

	Response Percent	Response Count
Yes.	**93.5%**	**72**
No.	6.5%	5
answered question		**77**
skipped question		**2**

6. If yes, what are those disadvantages? Please check all that apply.

	Response Percent	Response Count
Wasted effort in running control experiments.	22.5%	16
Lack of standardisation/impossible to compare results across papers.	**84.5%**	**60**
Some problems are "toy problems".	81.7%	58
Other (please specify)		22
answered question		**71**
skipped question		**8**

7. What forms of GP do you use? Please check all that apply.

		Response Percent	Response Count
Standard GP (i.e. tree-based GP).		**83.3%**	**65**
Cartesian GP.		11.5%	9
Linear GP.		25.6%	20
Grammatical GP (e.g. Grammatical Evolution).		38.5%	30
Stack-based GP (e.g. PushGP).		15.4%	12
Finite-state machine GP (e.g. Evolutionary Programming).		3.8%	3
Standard GP with strong typing or other modifications.		41.0%	32
Estimation of Distribution Algorithm GP.		12.8%	10
Generative and developmental representations.		17.9%	14
Evolution of neural networks (e.g. NEAT).		10.3%	8
Other (please specify)		14.1%	11
		answered question	**78**
		skipped question	**1**

8. Which form of GP do you use most often?

		Response Percent	Response Count
Standard GP (i.e. tree-based GP).		**35.9%**	**28**
Cartesian GP.		2.6%	2
Linear GP.		6.4%	5
Grammatical GP (e.g. Grammatical Evolution).		20.5%	16
Stack-based GP (e.g. PushGP).		7.7%	6
Finite-state machine GP (e.g. Evolutionary Programming).		0.0%	0
Standard GP with strong typing or other modifications.		15.4%	12
Estimation of Distribution Algorithm GP.		2.6%	2
Generative and developmental representations.		1.3%	1
Evolution of neural networks (e.g. NEAT).		0.0%	0
Other (please specify)		7.7%	6
		answered question	**78**
		skipped question	**1**

9. Have you heard of previous efforts at standardising GP benchmarks, e.g. GP-Beagle?

		Response Percent	Response Count
Yes.		**54.5%**	**42**
No.		45.5%	35
		answered question	**77**
		skipped question	**2**

10. Have you used previous GP benchmark suites?

		Response Percent	Response Count
Yes.		19.5%	15
No.		**72.7%**	**56**
Don't know.		7.8%	6
		answered question	**77**
		skipped question	**2**

11. Have you used benchmark suites in areas outside GP? (for example, in real-valued optimisation)

		Response Percent	Response Count
Yes.		46.8%	36
No.		**48.1%**	**37**
Don't know.		5.2%	4
		answered question	**77**
		skipped question	**2**

12. Are you in favour of the creation of a standardised GP benchmark suite?

		Response Percent	Response Count
Yes.		**83.1%**	**64**
No.		9.1%	7
Don't know.		7.8%	6
		answered question	**77**
		skipped question	**2**

13. If yes, why? Please check all that apply.

		Response Percent	Response Count
Allow easier comparisons between techniques.		**95.6%**	**65**
Standardisation.		75.0%	51
Demonstrate GP's success to outside parties.		48.5%	33
Find out which techniques are the best.		50.0%	34
Useful as a first step in evaluating a new technique		75.0%	51
Other (please specify)		7.4%	5
		answered question	**68**
		skipped question	**11**

14. If not, why not? Please check all that apply.

	Response Percent	Response Count
Existing benchmark practices are good enough.	0.0%	0
It might be a good idea, but the required community effort and support would not materialise.	13.6%	3
Researchers might concentrate on benchmarks, ignoring issues not covered in the benchmarks.	**86.4%**	**19**
Benchmarks are unrealistic.	18.2%	4
Incremental improvements on benchmarks aren't as important as fundamental improvements.	54.5%	12
Other (please specify)	31.8%	7
	answered question	**22**
	skipped question	**57**

15. If community consensus was in favour of creating a standardised benchmark suite, would you use it?

	Response Percent	Response Count
Yes.	**84.4%**	**65**
No.	3.9%	3
Don't know.	11.7%	9
	answered question	**77**
	skipped question	**2**

16. If yes, would you also use your own benchmark problems sometimes?

	Response Percent	Response Count
Would use standardised suite only.	2.9%	2
Would use both standardised suite and my preferred problems.	**91.4%**	**64**
Would use my preferred problems only.	1.4%	1
Don't know.	2.9%	2
Not applicable (e.g., I don't run experiments).	1.4%	1
answered question		**70**
skipped question		**9**

17. Should a benchmark suite aim to include real-world problems, synthetic problems, or a mixture of both?

	Response Percent	Response Count
Real-world.	2.6%	2
Synthetic.	0.0%	0
Both.	**93.4%**	**71**
Don't know.	1.3%	1
Other (please specify)	2.6%	2
answered question		**76**
skipped question		**3**

18. How important are these properties of benchmarks?

	Not Important				Very Important	Rating Average	Response Count
Fast in CPU time.	16.2% (12)	14.9% (11)	**32.4% (24)**	27.0% (20)	9.5% (7)	2.99	74
Easy to implement and run.	1.4% (1)	9.5% (7)	9.5% (7)	36.5% (27)	**43.2% (32)**	4.11	74
Realistically difficult.	0.0% (0)	2.7% (2)	14.9% (11)	**41.9% (31)**	40.5% (30)	4.20	74
Tunable, i.e. easier and harder versions available for testing scalability.	5.4% (4)	8.1% (6)	18.9% (14)	**37.8% (28)**	29.7% (22)	3.78	74
Representation-independent.	2.7% (2)	13.5% (10)	21.6% (16)	**31.1% (23)**	**31.1% (23)**	3.74	74
Open-source.	2.7% (2)	4.0% (3)	6.7% (5)	18.7% (14)	**68.0% (51)**	4.45	75
Reference implementation available.	0.0% (0)	4.1% (3)	18.9% (14)	16.2% (12)	**60.8% (45)**	4.34	74
					answered question		**75**
					skipped question		**4**

19. What application domains and problem types should the benchmark suite contain? Please check all that apply.

		Response Percent	Response Count
Symbolic regression.		**82.7%**	**62**
Boolean Functions.		60.0%	45
Route-finding/planning.		58.7%	44
Constructed problems.		53.3%	40
"True programming" (searching, sorting, object-oriented GP, etc.)		76.0%	57
Classification.		73.3%	55
Other (please specify)		34.7%	26
		answered question	**75**
		skipped question	**4**

20. If using a standardised benchmark suite, how tightly should details be specified? Please bear in mind that sometimes specifying details prevents comparison of non-standard or new techniques

	Should be open to change				Must be specified precisely	Rating Average	Response Count
Population size, number of generations, etc., or budget of fitness evaluations.	**35.6% (26)**	11.0% (8)	11.0% (8)	15.1% (11)	27.4% (20)	2.88	73
Crossover operators, mutation operators, initialisation method, etc.	**53.4% (39)**	6.8% (5)	9.6% (7)	13.7% (10)	16.4% (12)	2.33	73
Generation of random constants and similar, where applicable.	**38.4% (28)**	11.0% (8)	15.1% (11)	17.8% (13)	17.8% (13)	2.66	73
Function and terminal sets.	21.9% (16)	8.2% (6)	11.0% (8)	20.5% (15)	**38.4% (28)**	3.45	73
Variable ranges (e.g. x = 0.0, 0.1, ... 1.0).	19.2% (14)	5.5% (4)	12.3% (9)	19.2% (14)	**43.8% (32)**	3.63	73
Training and testing splits.	9.7% (7)	9.7% (7)	5.6% (4)	25.0% (18)	**50.0% (36)**	3.96	72
					answered question		**73**
					skipped question		**6**

21. Are there any other details which should be specified?

	Response Count
	33
answered question	**33**
skipped question	**46**

22. Can you suggest any existing benchmark problems you think SHOULD be part of a benchmark suite? Please give reasons if possible. Please supply enough information to precisely identify the problems.

	Response Count
	19
answered question	19
skipped question	60

23. Can you suggest any existing benchmark problems you think SHOULD NOT be part of a benchmark suite? Please give reasons if possible. Please supply enough information to precisely identify the problems.

	Response Count
	18
answered question	18
skipped question	61

24. Did you attend the GECCO 2012 talk "Genetic Programming Needs Better Benchmarks"?

	Response Percent	Response Count
Yes.	37.3%	28
No.	**62.7%**	**47**
answered question		75
skipped question		4

25. Have you read the GECCO 2012 paper "Genetic Programming Needs Better Benchmarks"?

		Response Percent	Response Count
Yes.		**57.3%**	**43**
No.		42.7%	32
		answered question	**75**
		skipped question	**4**

26. For how many years have you worked in or studied GP?

	Response Count
	75
answered question	**75**
skipped question	**4**

27. Please rate your knowledge and understanding of GP benchmark issues.

	Know very little				Expert	Rating Average	Response Count
	5.3% (4)	5.3% (4)	**37.3% (28)**	33.3% (25)	18.7% (14)	3.55	75
					answered question		**75**
					skipped question		**4**

28. Where did you hear about this questionaire?

		Response Percent	Response Count
"Genetic Programming Needs Better Benchmarks" Presentation at GECCO 2012		29.3%	22
Elsewhere at GECCO 2012		10.7%	8
By reading the "Genetic Programming Needs Better Benchmarks" paper.		4.0%	3
Through a Genetic Programming mailing list.		**37.3%**	**28**
Other (please specify)		18.7%	14
		answered question	**75**
		skipped question	**4**

29. Please give any other comments you wish to make.

	Response Count
	27
answered question	**27**
skipped question	**52**

30. If you would like to participate in further discussion, please provide your email address.

	Response Count
	37
answered question	**37**
skipped question	**42**