

Testausraportti

Testauksesta yleisesti

Tässä työssä käytettävät testausmenetelmät ovat yksikkötestaus ja suorituskkytestaus. Molempiin testaustapoihin on yritetty valita mahdollisimman kattava sarja testejä joilla voidaan

- a) Todentaa, että ohjelma toimii oikein
- b) On riittävän suorituskkyinen tämän kurssin vaatimustaso huomioiden

Yksikkötestaus

Yksikkötestien pääasiallinen tarkoitus on todentaa, että ohjelma toimii oikein. Ohjelma on testattu kahdella eri sarjalla yksikkötestejä. Toinen **<sat_unit_test.py>** sisältää testit, joista jokaiselle löytyy totuusjakauma (satisfiable) ja toinen **<unsat_unit_test.py>** joille ei löydy totuusjakaumaa (unsatisfiable).

Satisfiable sarjan testit on kopioitu CaDiCal -projektin repositorysta [1]. Hakemisto unit-tests/sat sisältää testitiedostojen (.cnf) lisäksi myös vastaavat .sol-tiedostot. Ohjelma ei käytä .sol-tiedostoja mihinkään, mutta ne on otettu mukaan todisteeksi siitä, että kyseiselle testille on ainakin yksi mahdollinen ratkaisu olemassa.

Unsatisfiable sarjan testejä on vähemmän ja ne löytyvät hakemistosta unit-tests/unsat. Testitapausten generointiin käytin ChatGPT:tä ja tarkistin oikeellisuuden netistä löytyvillä online solvereilla esim [2]

Alla testiajojen tulokset:

sat_unit_test.py

Executing satisfiable testset:

1. sat9.cnf
2. prime121.cnf
3. sat8.cnf
4. prime169.cnf
5. prime961.cnf
6. prime25.cnf
7. sat11.cnf
8. sat10.cnf
9. prime2209.cnf
10. sat12.cnf
11. prime9.cnf
12. sat13.cnf
13. prime4.cnf
14. prime361.cnf
15. prime1849.cnf

16. prime1681.cnf
17. empty.cnf
18. sat3.cnf
19. sat2.cnf
20. prime289.cnf
21. prime841.cnf
22. sat1.cnf
23. prime1369.cnf
24. prime529.cnf
25. sat4.cnf
26. sat6.cnf
27. sat7.cnf
28. prime49.cnf

Testing file: sat9.cnf

Satisfiable, execution time for sat9.cnf: 0.00005221367 seconds

Testing file: prime121.cnf

Satisfiable, execution time for prime121.cnf: 0.11610674858 seconds

Testing file: sat8.cnf

Satisfiable, execution time for sat8.cnf: 0.00002098083 seconds

Testing file: prime169.cnf

Satisfiable, execution time for prime169.cnf: 0.56152606010 seconds

Testing file: prime961.cnf

Satisfiable, execution time for prime961.cnf: 0.70828914642 seconds

Testing file: prime25.cnf

Satisfiable, execution time for prime25.cnf: 0.01471114159 seconds

Testing file: sat11.cnf

Satisfiable, execution time for sat11.cnf: 0.00001788139 seconds

Testing file: sat10.cnf

Satisfiable, execution time for sat10.cnf: 0.00001406670 seconds

Testing file: prime2209.cnf

Satisfiable, execution time for prime2209.cnf: 7.29488992691 seconds

Testing file: sat12.cnf

Satisfiable, execution time for sat12.cnf: 0.00003910065 seconds

Testing file: prime9.cnf

Satisfiable, execution time for prime9.cnf: 0.00184702873 seconds

Testing file: sat13.cnf

Satisfiable, execution time for sat13.cnf: 0.00001811981 seconds

Testing file: prime4.cnf

Satisfiable, execution time for prime4.cnf: 0.00012683868 seconds

Testing file: prime361.cnf

Satisfiable, execution time for prime361.cnf: 1.28371381760 seconds

Testing file: prime1849.cnf

Satisfiable, execution time for prime1849.cnf: 8.34920978546 seconds

Testing file: prime1681.cnf

Satisfiable, execution time for prime1681.cnf: 17.39548277855 seconds

Testing file: empty.cnf

Satisfiable, execution time for empty.cnf: 0.00000882149 seconds

Testing file: sat3.cnf

Satisfiable, execution time for sat3.cnf: 0.00001573563 seconds

Testing file: sat2.cnf

Satisfiable, execution time for sat2.cnf: 0.00000882149 seconds

Testing file: prime289.cnf

Satisfiable, execution time for prime289.cnf: 2.28475093842 seconds

Testing file: prime841.cnf

Satisfiable, execution time for prime841.cnf: 4.48117399216 seconds

Testing file: sat1.cnf

Satisfiable, execution time for sat1.cnf: 0.00001692772 seconds

Testing file: prime1369.cnf

Satisfiable, execution time for prime1369.cnf: 15.83293104172 seconds

Testing file: prime529.cnf

Satisfiable, execution time for prime529.cnf: 1.87660312653 seconds

Testing file: sat4.cnf

Satisfiable, execution time for sat4.cnf: 0.00001597404 seconds

Testing file: sat6.cnf

Satisfiable, execution time for sat6.cnf: 0.00001692772 seconds

Testing file: sat7.cnf

Satisfiable, execution time for sat7.cnf: 0.00001478195 seconds

Testing file: prime49.cnf

Satisfiable, execution time for prime49.cnf: 0.02297210693 seconds

Average execution time for satisfiable testset: 2.150878 seconds

unsat_unit_test.py

Executing unsatisfiable testset:

1. unsat4.cnf

2. unsat1.cnf

3. unsat2.cnf

4. unsat3.cnf

Testing file: unsat4.cnf

Unsatisfiable, execution time for unsat4.cnf: 0.00005602837 seconds

Testing file: unsat1.cnf

Unsatisfiable, execution time for unsat1.cnf: 0.00003504753 seconds

Testing file: unsat2.cnf

Unsatisfiable, execution time for unsat2.cnf: 0.00000214577 seconds

Testing file: unsat3.cnf

Unsatisfiable, execution time for unsat3.cnf: 0.00001001358 seconds

Average execution time for unsatisfiable testset: 0.000026 seconds

Suorituskykytestaus:

Kansiosta **perf-tests** löytyy suorituskykytestit, jotka on generoitu käyttäen [3]

Testien generointiin käytettiin komentoa **<cnfgen randkcnf 3 n m>**

Suorituskykytestauksessa käytetään 3-SAT ($n = 100$) testikokoelmia, jotka ovat lähellä vaihteensiirtymää (lausekkeiden ja muuttujien määrän suhde) $c \approx 4,25$,

missä $c = m / n$,

n = muuttujien määrä,

m = lausekkeiden määrä.

Kokeellisesti on osoitettu, että vaikeimmat tapaukset satunnaisessa 3-SAT-ongelmassa ovat ne, joissa lausekkeiden ja muuttujien määrän suhde on lähellä 4,25 [4]

Alla suorituskkytestauksen tulokset testisarjoittain:

Executing perftest testset n=100, m=420:

1. 1_100_420.cnf
2. 8_100_420.cnf
3. 4_100_420.cnf
4. 10_100_420.cnf
5. 7_100_420.cnf
6. 2_100_420.cnf
7. 5_100_420.cnf
8. 9_100_420.cnf
9. 3_100_420.cnf
10. 6_100_420.cnf

Testing file: 1_100_420.cnf

Unsatisfiable, execution time for 1_100_420.cnf: 0.57186102867 seconds

Testing file: 8_100_420.cnf

Satisfiable, execution time for 8_100_420.cnf: 0.35141992569 seconds

Testing file: 4_100_420.cnf

Satisfiable, execution time for 4_100_420.cnf: 0.12238311768 seconds

Testing file: 10_100_420.cnf

Unsatisfiable, execution time for 10_100_420.cnf: 0.94839501381 seconds

Testing file: 7_100_420.cnf

Satisfiable, execution time for 7_100_420.cnf: 0.02077889442 seconds

Testing file: 2_100_420.cnf

Unsatisfiable, execution time for 2_100_420.cnf: 0.89505028725 seconds

Testing file: 5_100_420.cnf

Satisfiable, execution time for 5_100_420.cnf: 0.05135512352 seconds

Testing file: 9_100_420.cnf

Unsatisfiable, execution time for 9_100_420.cnf: 1.09147119522 seconds

Testing file: 3_100_420.cnf

Unsatisfiable, execution time for 3_100_420.cnf: 0.64301085472 seconds

Testing file: 6_100_420.cnf

Satisfiable, execution time for 6_100_420.cnf: 0.13215112686 seconds

Average execution time for testset n=100, m=420: 0.482788 seconds

Executing perfest testset n=100, m=428:

1. 3_100_428.cnf
2. 6_100_428.cnf
3. 5_100_428.cnf
4. 9_100_428.cnf
5. 7_100_428.cnf
6. 2_100_428.cnf
7. 1_100_428.cnf
8. 8_100_428.cnf
9. 10_100_428.cnf
10. 4_100_428.cnf

Testing file: 3_100_428.cnf

Satisfiable, execution time for 3_100_428.cnf: 0.01887512207 seconds

Testing file: 6_100_428.cnf

Satisfiable, execution time for 6_100_428.cnf: 1.31972813606 seconds

Testing file: 5_100_428.cnf

Satisfiable, execution time for 5_100_428.cnf: 0.30886077881 seconds

Testing file: 9_100_428.cnf

Unsatisfiable, execution time for 9_100_428.cnf: 0.75980806351 seconds

Testing file: 7_100_428.cnf

Unsatisfiable, execution time for 7_100_428.cnf: 0.63043618202 seconds

Testing file: 2_100_428.cnf

Satisfiable, execution time for 2_100_428.cnf: 0.67050194740 seconds

Testing file: 1_100_428.cnf

Satisfiable, execution time for 1_100_428.cnf: 0.80829000473 seconds

Testing file: 8_100_428.cnf

Satisfiable, execution time for 8_100_428.cnf: 0.06796002388 seconds

Testing file: 10_100_428.cnf

Satisfiable, execution time for 10_100_428.cnf: 0.19106030464 seconds

Testing file: 4_100_428.cnf

Unsatisfiable, execution time for 4_100_428.cnf: 0.78174018860 seconds

Average execution time for testset n=100, m=428: 0.555726 seconds

Executing perfest testset n=100, m=429:

1. 3_100_429.cnf
2. 6_100_429.cnf
3. 5_100_429.cnf
4. 9_100_429.cnf
5. 7_100_429.cnf
6. 2_100_429.cnf
7. 1_100_429.cnf
8. 8_100_429.cnf
9. 4_100_429.cnf
10. 10_100_429.cnf

Testing file: 3_100_429.cnf

Satisfiable, execution time for 3_100_429.cnf: 0.14326405525 seconds

Testing file: 6_100_429.cnf

Unsatisfiable, execution time for 6_100_429.cnf: 0.50987386703 seconds

Testing file: 5_100_429.cnf

Satisfiable, execution time for 5_100_429.cnf: 0.19149208069 seconds

Testing file: 9_100_429.cnf

Satisfiable, execution time for 9_100_429.cnf: 0.12090301514 seconds

Testing file: 7_100_429.cnf

Unsatisfiable, execution time for 7_100_429.cnf: 0.75366616249 seconds

Testing file: 2_100_429.cnf

Unsatisfiable, execution time for 2_100_429.cnf: 0.70695662498 seconds

Testing file: 1_100_429.cnf

Satisfiable, execution time for 1_100_429.cnf: 0.14005494118 seconds

Testing file: 8_100_429.cnf

Unsatisfiable, execution time for 8_100_429.cnf: 1.26583981514 seconds

Testing file: 4_100_429.cnf

Satisfiable, execution time for 4_100_429.cnf: 0.33482813835 seconds

Testing file: 10_100_429.cnf

Satisfiable, execution time for 10_100_429.cnf: 0.01224493980 seconds

Average execution time for testset n=100, m=429: 0.417912 seconds

Executing perfest testset n=100, m=450:

1. 1_3_100_450.cnf

2. 7_100_450.cnf

3. 2_3_100_450.cnf

4. 10_100_450.cnf

5. 4_100_450.cnf

6. 8_100_450.cnf

7. 6_100_450.cnf

8. 3_100_450.cnf

9. 9_100_450.cnf

10. 5_100_450.cnf

Testing file: 1_3_100_450.cnf

Unsatisfiable, execution time for 1_3_100_450.cnf: 0.95497393608 seconds

Testing file: 7_100_450.cnf

Unsatisfiable, execution time for 7_100_450.cnf: 1.09054207802 seconds

Testing file: 2_3_100_450.cnf

Satisfiable, execution time for 2_3_100_450.cnf: 0.10038781166 seconds

Testing file: 10_100_450.cnf

Unsatisfiable, execution time for 10_100_450.cnf: 0.51843667030 seconds

Testing file: 4_100_450.cnf

Unsatisfiable, execution time for 4_100_450.cnf: 0.54082012177 seconds

Testing file: 8_100_450.cnf

Satisfiable, execution time for 8_100_450.cnf: 0.16631770134 seconds

Testing file: 6_100_450.cnf

Unsatisfiable, execution time for 6_100_450.cnf: 0.72718310356 seconds

Testing file: 3_100_450.cnf

Unsatisfiable, execution time for 3_100_450.cnf: 0.78898096085 seconds

Testing file: 9_100_450.cnf

Unsatisfiable, execution time for 9_100_450.cnf: 0.82096910477 seconds

Testing file: 5_100_450.cnf

Unsatisfiable, execution time for 5_100_450.cnf: 0.81609177589 seconds

Average execution time for testset n=100, m=450: 0.652470 seconds

Executing perf test testset n=100, m=480:

1. 7_100_480.cnf
2. 2_100_480.cnf
3. 1_100_480.cnf
4. 8_100_480.cnf
5. 4_100_480.cnf
6. 10_100_480.cnf
7. 3_100_480.cnf
8. 6_100_480.cnf
9. 5_100_480.cnf
10. 9_100_480.cnf

Testing file: 7_100_480.cnf

Unsatisfiable, execution time for 7_100_480.cnf: 0.63577604294 seconds

Testing file: 2_100_480.cnf

Unsatisfiable, execution time for 2_100_480.cnf: 0.38602495193 seconds

Testing file: 1_100_480.cnf

Unsatisfiable, execution time for 1_100_480.cnf: 0.60872602463 seconds

Testing file: 8_100_480.cnf

Unsatisfiable, execution time for 8_100_480.cnf: 0.49737286568 seconds

Testing file: 4_100_480.cnf

Unsatisfiable, execution time for 4_100_480.cnf: 0.77195477486 seconds

Testing file: 10_100_480.cnf

Unsatisfiable, execution time for 10_100_480.cnf: 0.49034023285 seconds

Testing file: 3_100_480.cnf

Unsatisfiable, execution time for 3_100_480.cnf: 0.49321508408 seconds

Testing file: 6_100_480.cnf

Unsatisfiable, execution time for 6_100_480.cnf: 0.52523422241 seconds

Testing file: 5_100_480.cnf

Unsatisfiable, execution time for 5_100_480.cnf: 0.53323316574 seconds

Testing file: 9_100_480.cnf

Satisfiable, execution time for 9_100_480.cnf: 0.63478398323 seconds

Average execution time for testset n=100, m=480: 0.557666 seconds

Lähdeluettelo:

[1] <https://github.com/arminbiere/cadical>.

[2] <https://jgalenson.github.io/research.js/demos/minisat.html>

[3] <https://massimolauria.net/cnfgcn>

[4] <https://dl.acm.org/doi/10.5555/2832249.2832300>