

STS Test Results

Moderators: H.G.Muller, Harvey Williamson, bob

Post Reply

🔧

▼

Search this topic...

🔍

⚙️

14 posts

1

2

>

MikeB

“

STS Test Results

📅 Sun Jun 14, 2015 1:47 am

Latest developmental version of crafty.

At 10 seconds per move. iMac Retina 5K , 4 CORE i7 4790 4.0Ghz

CODE: SELECT ALL

```
test results summary:

total positions searched.....      1300
number right.....                  1056
number wrong.....                   244
percentage right.....                81
percentage wrong.....                18
total nodes searched.....355918365972
average search depth.....           24.6
nodes per second.....               27160114
total time.....                     218:24
```

not sure how this stacks against other programs...

⬆️

Ferdy

“

Re: STS Test Results

📅 Sun Jun 14, 2015 4:19 am

“ MikeB wrote:

Latest developmental version of crafty.

At 10 seconds per move. iMac Retina 5K , 4 CORE i7 4790 4.0Ghz

CODE: SELECT ALL

```
test results summary:

total positions searched.....      1300
number right.....                  1056
number wrong.....                   244
percentage right.....                81
percentage wrong.....                18
```

```
total nodes searched.....355918365972
average search depth.....      24.6
nodes per second.....      27160114
total time.....      218:24
```

not sure how this stacks against other programs...

STS has 15 themes now equivalent to $100 \times 15 = 1500$ positions.

Download the complete epd here.

<https://sites.google.com/site/strategictestsuite/>

or if you process the alternative moves and point system, download the complete 15 themes in the following, at the bottom of the first post. There are couple of changes in the alternative moves, and also uniform formatting in the epd with additional opcodes converting SAN to LAN.

<http://www.talkchess.com/forum/viewtopic.php?p=56653>

BTW is there an option for Crafty to output move in LAN format like e2e4 instead of e4?

MikeB



Re: STS Test Results

Sun Jun 14, 2015 4:15 pm

“ Ferdy wrote:

“ MikeB wrote:

Latest developmental version of crafty.

At 10 seconds per move. iMac Retina 5K , 4 CORE i7 4790 4.0Ghz

CODE: [SELECT ALL](#)

test results summary:

```
total positions searched.....      1300
number right.....      1056
number wrong.....      244
percentage right.....      81
percentage wrong.....      18
total nodes searched.....355918365972
average search depth.....      24.6
nodes per second.....      27160114
total time.....      218:24
```

not sure how this stacks against other programs...

STS has 15 themes now equivalent to $100 \times 15 = 1500$ positions.

Download the complete epd here.

<https://sites.google.com/site/strategictestsuite/>

or if you process the alternative moves and point system, download the complete 15 themes in the following, at the bottom of the first post. There are couple of changes in the alternative moves, and also uniform formatting in the epd with additional opcodes converting SAN to LAN.

<http://www.talkchess.com/forum/viewtopic.php?p=56653>

BTW is there an option for Crafty to output move in LAN format like e2e4 instead of e4?

Appreciate the reply. Crafty does output in LAN – option "output long". Looks like your program is pc. If you are interested, I can try to compile a mac version.

st 200ms– single core:

CODE: [SELECT ALL](#)

test results summary:

total positions searched.....	1500
number right.....	862
number wrong.....	638
percentage right.....	57
percentage wrong.....	42
total nodes searched.....	2055634974
average search depth.....	14.0
nodes per second.....	6852116
total time.....	5:00

White(1):

st 200ms – 4 core:

CODE: [SELECT ALL](#)

test results summary:

total positions searched.....	1500
number right.....	953
number wrong.....	547
percentage right.....	63
percentage wrong.....	36
total nodes searched.....	6717873153
average search depth.....	15.8
nodes per second.....	22392910
total time.....	5:00

White(1):

MikeB



Re: STS Test Results

Sun Jun 14, 2015 5:05 pm

st 800ms single core:

CODE: [SELECT ALL](#)

test results summary:

total positions searched.....	1500
number right.....	979
number wrong.....	521
percentage right.....	65
percentage wrong.....	34
total nodes searched.....	8627711134
average search depth.....	16.7
nodes per second.....	7189759
total time.....	20:00

White(1):

st 800ms 4 core

CODE: [SELECT ALL](#)

test results summary:

```

total positions searched..... 1500
number right..... 1043
number wrong..... 457
percentage right..... 69
percentage wrong..... 30
total nodes searched..... 30004379169
average search depth..... 18.9
nodes per second..... 25003649
total time..... 20:00
White(1):

```

Ferdy



Re: STS Test Results

Mon Jun 15, 2015 6:27 am

MikeB wrote:

Ferdy wrote:

MikeB wrote:

Latest developmental version of crafty.

At 10 seconds per move. iMac Retina 5K , 4 CORE i7 4790 4.0Ghz

CODE: SELECT ALL

test results summary:

```

total positions searched..... 1300
number right..... 1056
number wrong..... 244
percentage right..... 81
percentage wrong..... 18
total nodes searched..... 355918365972
average search depth..... 24.6
nodes per second..... 27160114
total time..... 218:24

```

not sure how this stacks against other programs...

STS has 15 themes now equivalent to $100 \times 15 = 1500$ positions.

Download the complete epd here.

<https://sites.google.com/site/strategictestsuite/>

or if you process the alternative moves and point system, download the complete 15 themes in the following, at the bottom of the first post. There are couple of changes in the alternative moves, and also uniform formatting in the epd with additional opcodes converting SAN to LAN.

[http://www.talkchess.com/forum/viewtopic.php ... 06&t=56653](http://www.talkchess.com/forum/viewtopic.php?t=56653)

BTW is there an option for Crafty to output move in LAN format like e2e4 instead of e4?

Appreciate the reply. Crafty does output in LAN – option "output long".

I just knew it, unfortunately it also outputs the Piece letter as in Ng1f3. I thought it is g1f3, as in uci output. So in this case the tool is of no use to calculate points. Later I will implement SAN, but I need to create a new

format on the epd. Also on my list is to support st (decimal and integer values in seconds) command for WB engines.

“

Looks like your program is pc. If you are interested, I can try to compile a mac version.

I am using windows, the source is a python script and converted to exe file. I am using python 2.7.6. Does python run on your system?

“

st 200ms- single core:

CODE: [SELECT ALL](#)

test results summary:

```
total positions searched..... 1500
number right..... 862
number wrong..... 638
percentage right..... 57
percentage wrong..... 42
total nodes searched..... 2055634974
average search depth..... 14.0
nodes per second..... 6852116
total time..... 5:00
White(1):
```

I run Crafty on my system also at 200ms/pos.

CODE: [SELECT ALL](#)

Raw nodes per second: 5242487
Total elapsed time: 37.05

STS 1-15, 0.2s/pos

test results summary:

```
total positions searched..... 1500
number right..... 826
number wrong..... 674
percentage right..... 55
percentage wrong..... 44
total nodes searched..... 1055659387
average search depth..... 12.9
nodes per second..... 3468114
```

Your machine is faster than mine. Here are some results from different engines, also at 200ms. We can only compare the BestCnt line.

CODE: [SELECT ALL](#)

BestCnt	01	02	03	04	05	06	07	08	09	10	11	12	13
Score	697	682	721	679	711	765	683	585	632	767	591	719	744
Score(%)	69.7	68.2	72.1	67.9	71.1	76.5	68.3	58.5	63.2	76.7	59.1	71.9	74.4

Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz

Engine: Stockfish 6 64 POPCNT

Hash: 128, Threads: 1, time/pos: 0.200s

Test duration: 00:05:20

Expected time to finish: 00:05:45

STS ID	STS1	STS2	STS3	STS4	STS5	STS6	STS7	STS8	STS9	STS10	STS11	STS12	STS13
NumPos	100	100	100	100	100	100	100	100	100	100	100	100	100
BestCnt	84	76	73	70	77	76	69	68	69	81	73	72	78
Score	874	847	818	807	821	893	787	794	765	864	795	786	842

Comparison based on my machine. Rating in parenthesis is from CCRL 40/4.

CODE: [SELECT ALL](#)

Arasan 17.5 770/1500 (2849)
 Rodent 1.7 build 1 824/1500 (2833)
 Crafty v24.1 (1 cpus) 826/1500 (No record, but v24.0 is 2801)
 Rhetoric 1.4.1 x64 857/1500 (2794)
 Deuterium v14.3.34.130 859/1500 (2889)
 Nemo SP64o 1.0.1 Beta 865/1500 (2864)
 Gaviota v1.0 884/1500 (2895)
 Senpai 1.0 941/1500 (3016)
 Stockfish 6 64 POPCNT 1089/1500 (3317)

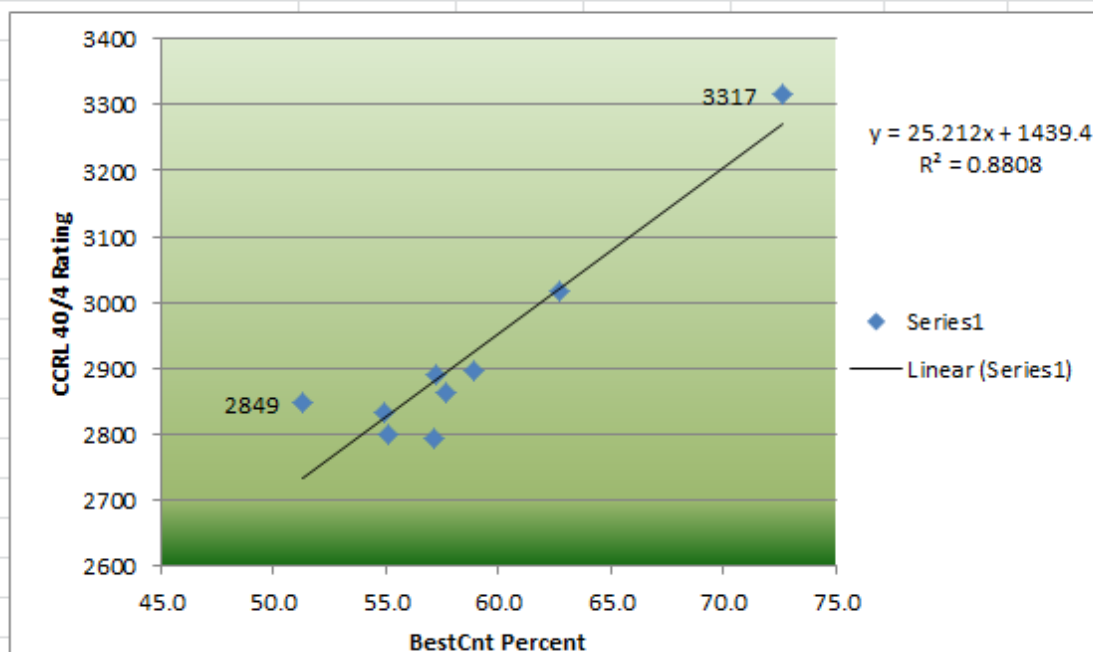
Not really a good fit, just a rough formula to get a rating estimate for CCRL 40/4 based on bestmove counts percentage from 1500 pos, at 200ms/pos.

I think Crafty does fine here as with other engines though the data is not that big.

CODE: [SELECT ALL](#)

$$\text{sts_rating} = 25.212 \times \text{bestmove_percentage} + 1439.4$$

Engines at 200ms/pos	CCRL 40/4 Rating	BestCnt	Max BestCnt	Rate(%)	STS Rating	Diff (ccrl-sts)
Arasan 17.5	2849	770	1500	51.3	2734	115
Rodent 1.7 build 1	2833	824	1500	54.9	2824	9
Crafty v24.1 (1 cpus)	2801	826	1500	55.1	2828	-27
Rhetoric 1.4.1 x64	2794	857	1500	57.1	2880	-86
Deuterium v14.3.34.130	2889	859	1500	57.3	2883	6
Nemo SP64o 1.0.1 Beta	2864	865	1500	57.7	2893	-29
Gaviota v1.0	2895	884	1500	58.9	2925	-30
Senpai 1.0	3016	941	1500	62.7	3021	-5
Stockfish 6 64 POPCNT	3317	1089	1500	72.6	3270	47



MikeB



Re: STS Test Results

Wed Jun 17, 2015 1:51 am

“

...the source is a python script and converted to exe file. I am using python 2.7.6. Does python run on your system?

Yes – python scripts run on Mac. I can adjust output to LAN as well (with no piece indicators).



Ferdy

“

Re: STS Test Results

Wed Jun 17, 2015 6:24 am

“ MikeB wrote:

“

...the source is a python script and converted to exe file. I am using python 2.7.6. Does python run on your system?

Yes – python scripts run on Mac. I can adjust output to LAN as well (with no piece indicators).

I have implemented the SAN for WB engine, also support for st command both integer and with decimal number in sec, as in --st 0.2 --san options.

The link to script and new epd format is sent in PM. For WB engine automatic sts rating calculation is still not implemented, but you can get the points percentage.



MikeB

“

Re: STS Test Results

Sat Jul 18, 2015 2:41 pm

“ MikeB wrote:

CODE: [SELECT ALL](#)

test results summary:

total positions searched.....	1500
number right.....	1043
number wrong.....	457
percentage right.....	69
percentage wrong.....	30
total nodes searched.....	30004379169
average search depth.....	18.9
nodes per second.....	25003649
total time.....	20:00

White(1):

latest developmental version

CODE: [SELECT ALL](#)

test results summary:

total positions searched.....	1500
number right.....	1057
number wrong.....	443
percentage right.....	70
percentage wrong.....	29
total nodes searched.....	26303312891

```
average search depth..... 19.2
nodes per second..... 21918879
total time..... 20:00
White(1):
```

a tad slower and a tad more smarter

MikeB



Re: STS Test Results

Sat Jul 18, 2015 3:21 pm

MikeB wrote:

CODE: SELECT ALL

test results summary:

```
total positions searched..... 1500
number right..... 862
number wrong..... 638
percentage right..... 57
percentage wrong..... 42
total nodes searched..... 2055634974
average search depth..... 14.0
nodes per second..... 6852116
total time..... 5:00
White(1):
```

]

CODE: SELECT ALL

latest developmental version:

test results summary:

```
total positions searched..... 1500
number right..... 887
number wrong..... 613
percentage right..... 59
percentage wrong..... 40
total nodes searched..... 2165113989
average search depth..... 14.4
nodes per second..... 7217046
total time..... 5:00
```

User avatar xr_a_y



Re: STS Test Results

Mon Apr 23, 2018 4:16 pm

Wow just found this old STS results, so I decided to score Weini at this ... well not a good news, at 5sec per position (mean depth was around 11), Weini only gets 688/1500 right ...

Splitting that by STS pack this gives :

- 1: 47 (nothing special is implemented)
- 2: 36 (Wow ! I though my thing on this subject was working ... :-())
- 3: 49 (just PSQT)
- 4: 46 (just something via PSQT)

5: 51 (more or less same as in CPW)
6: 57
7: 37 (a king of tradeoff pieces bonus for the winning side is activated)
8: 38 (ok pawn storm was not activated in eval ...)
9: 45 (ok pawn storm was not activated in eval ...)
10: 51
11: 44 (Weini has both some king centralization via PSQT and king safety by keeping pawn in front of king, but king troppism was not activated for this test)
12: 46 (center control by threat look up was not activated, but PSQT have some stuff on this subject)
13: 65
14: 52 (just by PSQT)
15: 26 (... how can I defined "queen pointless exchange" in eval ?)

A lot of work to do ... Is this a good idea to work on evaluation based on STS ?

[Post Reply](#)

14 posts

1

2

[Return to "Computer Chess Club: General Topics"](#)[Jump to](#)[ChessUSA - Your Move Chess & Games](#) < [Board index](#)[The team](#)[Delete all board cookies](#)

All times are UTC

[Contact us](#)Powered by [phpBB®](#) Forum Software © phpBB Limited