

Jahresrückblick CSS-Rangliste

Tests und Statistik

Vorher etwas Statistik. Getestet wurden im letzten Jahr insgesamt 32 Engines, dabei wurden über 14.000 Partien absolviert. Dies fand immer unter den gleichen Bedingungen statt, sodass die Möglichkeit besteht, detaillierte Vergleiche zu ziehen. Unter den 32 Testkandidaten waren im vergangenen Jahr gleich sieben Engines, welche bisher noch nicht in der CSS-Rangliste vertreten waren. Von denen schafften mit Patriot v1.1.0, Thinker 4.6b und zumindest für kurze Zeit auch Green Light Chess 3.00.3.4 drei Engines die Qualifikation in die CSS-Rangliste. Mit Ruffian, Ktulu und Patriot erhielten die kommerziellen Schachprogramme einen ungewohnt hohen Zulauf. Aber es gab auch den umgekehrten Fall. Ed Schroeder aus den Niederlanden startete, aufbauend auf die starke Rebel 12-Engine, mit Pro Deo ein Freeware-Projekt.

Testkandidat	Adapter		Partien	Rating (Elo)		
				Vorgänger	aktuell	+/-
Ruffian 2.0.0	Kommerz.	UCI	480	2595	2649	54
Shredder 8	Kommerz.	Native	480	2744	2766	22
SOS 4 for Arena	Freeware	UCI	480	2577	2607	30
Crafty 19.08	Freeware	Native	480	2482	2543	61
AnMon 5.30	Freeware	UCI	500	2438	2490	52
Anaconda 1.6.2	Freeware	Native	480	2503	2523	20
Tao 5.6	Freeware	Native	480	2485	2526	41
Delfi 4.4	Freeware	WB	480	2520	2568	48
Patriot v1.1.0	Kommerz.	UCI	500		2502	
Amyan 1.592	Freeware	UCI	480	2505	2532	27
Aristarch 4.50	Freeware	UCI	480	2577	2621	44
Deep Sjeng 1.6	Kommerz.	UCI	480	2575	2604	29
Ktulu 5.1	Kommerz.	UCI	480	2500	2556	56
Thinker 4.6b	Freeware	WB	500		2576	
Little Goliath Revival 1.0.13	Freeware	UCI	250	2449	2486	37
WildCat 4	Freeware	WB	500		2504	
SmarThink 0.17a	Freeware	WB	480	2546	2600	54
Patriot 1.3.0	Kommerz.	UCI	480	2502	2540	38
Pro Deo 1.0	Freeware	WB	500	2578	2621	43
Jonny 2.64	Freeware	WB	160		2468	
Green Light Chess 3.00.3.4	Freeware	UCI	500		2515	
Quark 2.35 Paderborn	Freeware	WB	500		2500	
The King 3.33	Kommerz.	WB	480	2674	2657	-17
Little Goliath Revival	Freeware	UCI	500	2486	2536	50
Pharaon 3.1	Freeware	UCI	500	2478	2565	87
AnMon 5.40	Freeware	UCI	168	2490	2527	37
AnMon 5.50	Freeware	UCI	500	2527	2549	22
The Baron 1.5.0	Freeware	WB	168		2500	
Fritz 8 Bilbao	Kommerz.	Native	480	2720	2708	-12
Gandalf 6.0	Kommerz.	UCI	480	2583	2660	77
Amyan 1.594	Freeware	UCI	480	2532	2517	-15
Junior 9	Kommerz.	UCI	480	2658	2702	44
Summe			14386			37,2
Zusammenfassung - Vergleich Freeware und kommerzielle Engines						
	Anzahl	Neu	Partien	Elo i.M.	+/-	
Freeware Engines	22	6	9566	2540	40	
Kommerzielle Engines	10	1	4820	2634	32	

CSS-Rangliste -Stand Dezember 2003				
Rang	Schachengine	Score	Elo Gegner	Elo
1	Shredder 7.04	74,17%	2570	2752

2	Deep Fritz 8	70,42%	2571	2721
3	Chess Tiger 15.0	68,75%	2572	2707
4	Hiarcs 9	68,54%	2572	2706
5	Junior 8	62,92%	2573	2664
6	The King 3.23	62,50%	2574	2661
7	List 5.12	60,63%	2574	2648
8	Ruffian 1.0.1	54,58%	2576	2607
9	Gandalf 5.1	51,25%	2577	2584
10	Deep Sjeng 1.5	50,52%	2577	2579
11	Aristarch 4.21	50,42%	2577	2579
12	Rebel 12	48,54%	2578	2566
13	SOS.3 for Arena	48,33%	2578	2565
14	Nimzo 8	46,88%	2578	2555
15	SmarThink 0.16b	43,75%	2579	2534
16	El Chinito 3.1c0	42,40%	2579	2525
17	Pepito v1.59	40,63%	2580	2513
18	Amyan 1.59	40,42%	2580	2511
19	Anaconda 1.0	40,00%	2580	2508
20	Delfi 4.1	39,79%	2580	2507
21	Ktulu 4.0	37,60%	2581	2492
22	Crafty 19.01	37,50%	2581	2491
23	Yace Paderborn	36,88%	2581	2486
24	Tao 5.4	36,77%	2581	2486
25	Pharaon 2.62	35,83%	2581	2479

Schauen wir uns anhand der Abbildung 1 die in diesem Jahr durchgeführten Tests an. Wie aus der Grafik deutlich wird, konnten sich die Testkandidaten im Durchschnitt um erfreuliche 37 Elopunkte steigern. Das ist ein ausgezeichneter Wert, Hauptanteil an dieser Steigerung hatten die Freeware-Engines mit einer durchschnittlichen Steigerung von 40 Elopunkten. Die Profikollegen kamen auf einen Durchschnittswert von +32 Elo. Daran wird schon deutlich, dass die frei verfügbaren Engines in diesem Jahr aufgeholt haben. Um die nachfolgenden Angaben mit den Werten aus dem vergangenen Jahr besser vergleichen zu können, habe ich die Rangliste mit Stand von Dezember 2003 auszugsweise mit aufgeführt.

Jahresvergleich				
	Stand Dez. 2001	Stand Dez. 2002	Stand Dez. 2003	Stand Dez. 2004
Anzahl Engines	18	18	25	25
Durchschnittliche Spielstärke	2466	2566	2577	2609
Rating Rang 1	2715	2726	2752	2767
Rating Rang 25	2156	2472	2479	2521
Differenz Rang 1 zu Rang 25	559	254	273	246
Remisquote	18,82%	24,58%	25,45%	28,70%

Wie aus der Abbildung 2 zu erkennen ist, waren die in der CSS-Rangliste platzierten Engines durchschnittlich um 32 Elopunkte stärker. Hauptverantwortlich hierfür war die Begrenzung der Rangliste auf maximal 25 Engines. Dadurch wurde die Einstiegs-Elozahl durch die Qualifikation neuer starker Engines kontinuierlich nach oben gefahren. Während sich die Steigerung an der Tabellenspitze eher moderat verhielt, konnte am Tabellenende eine Elosteigerung von über 40 Elopunkten erzielt werden. Das gestiegene Gesamtniveau der Rangliste und der somit besseren Ausgeglichenheit der Gegner spiegelt sich im deutlich verminderten Elo-Unterschied zwischen Platz 1 und 25 der Rangliste wieder. Der Abstand konnte seit 2001 erstmalig deutlich gesenkt werden. Auch der gestiegene Remisfaktor zeugt von der besseren Ausgeglichenheit der Rangliste.

Vergleich der kommerziellen Engines				
	Stand Dez. 2001	Stand Dez. 2002	Stand Dez. 2003	Stand Dez. 2004
Anzahl Engines	7	8	10	12
Durchschnittliche Spielstärke	2590	2636	2650	2646
Beste Engine	2715	2726	2752	2767
Schlechteste Engine	2537	2542	2555	2534

Diff. beste zu schlechteste Engine	178	184	197	233
------------------------------------	-----	-----	-----	-----

Schaut man sich die Spielstärkeveränderungen bei den kommerziellen Programmen an, fällt auf, dass dank Shredder 8 die Spielstärke an der Tabellenspitze sogar noch wuchs, die durchschnittliche Spielstärke allerdings geringfügig nachließ. Das liegt aber eher daran, dass mit Patriot und Ktulu zwei Engines ins Profilager wechselten, welche noch nicht deren durchschnittliche Spielstärke erreichen. Für ordentlichen Zuwachs sorgte jedenfalls Gandalf 6.0, welcher allein für eine Elosteigerung von etwa 80 Punkten sorgte. Junior 9 legte immerhin noch 44 Punkte zu, während es bei Fritz 8 Bilbao und The King 3.33 nicht vorwärts ging. Beim Letztgenannten muss ich aber fairerweise erwähnen, dass die Engine erstmalig mit den etwas schwächeren "Default-Einstellungen" getestet wurde. Unter den früheren Testbedingungen (sel=12) dürfte die Verbesserung etwa bei 20–30 Elo liegen.

Vergleich der Freeware-Engines				
	Stand Dez. 2001	Stand Dez. 2002	Stand Dez. 2003	Stand Dez. 2004
Anzahl Engines	11	10	15	13
Durchschnittliche Spielstärke	2387	2510	2529	2575
Beste Engine	2494	2603	2648	2648
Schlechteste Engine	2156	2472	2479	2521
Diff. beste zu schlechteste Engine	338	131	169	127

Ganz anders sieht es im Amateurlager aus. Hier wurde durch starke Zugewinne wie von Pro Deo und Thinker sowie hervorragende Spielstärkeresteigerungen von Pharaon und Delfi einiges bewegt. Während die Freewarespitze stagniert (was aber daran liegt, dass noch niemand der Amateurliegen das außerordentliche Niveau von List überbietet), waren im unteren Spielstärkebereich erhebliche Steigerungen zu erkennen. Die Elozahl der schlechtesten Amateurengine konnte ebenso wie die durchschnittliche Spielstärke um 46 Punkte verbessert werden.

Unterschiede der kommerziellen Engines zu Freeware-Engines				
	Stand Dez. 2001	Stand Dez. 2002	Stand Dez. 2003	Stand Dez. 2004
Durchschnittliche Spielstärke	203	126	121	71
Beste Engine	221	123	104	119
Schlechteste Engine	381	70	76	13

Im direkten Vergleich wird deutlich, dass es durch die nach wie vor unangefochtenen Spitzenpositionen von Shredder und List kaum Veränderungen gab. Ja, hier konnte sich Shredder sogar noch etwas absetzen. Insgesamt betrachtet konnten die Amateure den Abstand zu den Profis um stolze 50 Elopunkte reduzieren. Die durchschnittliche Spielstärke weicht nur noch um 71 Elopunkte voneinander ab. Wie die Entwicklungen vorangeschritten sind, lässt sich auch ausgezeichnet am Vergleich mit Oldie Nimzo 8 anstellen. Dies ist die einzige Engine, welche seit Beginn der Rangliste vertreten ist.

Nimzo 8-Schwelle				
	Stand Dez. 2001	Stand Dez. 2002	Stand Dez. 2003	Stand Dez. 2004
Elo Rating von Nimzo 8	2547	2542	2555	2568
Gegnerische Spielstärke	2461	2567	2578	2611
Score von Nimzo 8	62,06%	46,47%	46,88%	43,96%
Ranglistenplatzierung	5	10	14	16
Eloabstand zum Tabellenersten	168	184	197	199
Eloabstand zum letzten Rang	391	70	76	47

Während Nimzo 8 durch die Steigerung des Ranglistenniveaus seine Elozahl selbst auch etwas steigern konnte, ist der Abstand zum Tabellenende mit 47 Elopunkten verdächtig nahe gerückt. Wenn die Entwicklung im Freeware-Bereich so fortschreitet, dauert es nicht mehr lange, bis die älteste noch aktuelle Profi-Engine aus der Rangliste fällt. Allerdings glaube ich nicht, dass dies schon in diesem Jahr passiert. Zur Tabellenspitze ist der Abstand von Nimzo 8 in diesem Jahr um fast 200 Elopunkte gewachsen. Auch daran zeigt sich, dass die Spielstärke trotz des extrem hohen Spielniveaus immer noch verbessert werden kann.

Ausblick

Ich bin schon sehr gespannt, was das neue Jahr uns Computerschachfreaks bringt. Gleich zu Beginn des neuen Jahres kam vom Ranglistenführer Shredder eine neue Version auf den Markt, welche die schon vorhandene Ausnahmestellung durch eine weitere Spielstärkesteigerung ausbauen konnte. Fritz-Autor Frans Morsch legte seine Prioritäten bei der Optimierung des Spielstiles mittlerweile auf das Spiel gegen menschliche Großmeister. Das Ergebnis war ein attraktiver Spielstil, der allerdings gegen die Computerkonkurrenz keine besseren Ergebnisse brachte.

Viel Bewegung erwarte ich auch bei den Jungprofis und denjenigen, die es werden wollen. So kündigte der russische Autor Sergej Markoff (SmarThink) eine kommerzielle UCI-Engine an. Auch bei Rahman Paidar (Ktulu) und Viktor Yelin (Patriot) sehe ich noch Verbesserungspotenzial. Den stärksten Aufwind hatten im vergangenen Jahr die Amateurprogrammierer. Sollte auch in diesem Jahr eine ähnliche Steigerung erfolgen, wird der Abstand zu den kommerziellen Konkurrenten verdächtig eng. Bei einer schon passablen Spielstärke sind weitere Verbesserungen oft nur in kleineren Schritten möglich, sodass ich hier eher eine Stagnation erwarte. Aber vielleicht werde ich hier auch eines Besseren belehrt. Ich werde dies durch die kommenden Tests interessiert beobachten und entsprechend gewohnt detailliert kommentieren. *(Klaus Wlotzka)*

Informationen zum Autor:

Klaus Wlotzka
