

ДЛЯ ТЕХ, КТО СКАЧИВАЛ МЕГАБАЗУ И ДОПОЛНЕНИЯ РАНЕЕ

Предлагается дополнение Унифицированной Мегабазы за 2018 год, 388420 партий, можно скачать по ссылке

<https://yadi.sk/d/21CNPXkSlm9pyA>

цена вопроса 300 рублей, как обычно, кошелек Вебмани R282782512403, потом вышлю пароль на архив.

Если есть желание (или для тех, кто скачает ВПЕРВЫЕ), можете ВМЕСТО дополнений сразу скачать за 450 рублей уже готовую Унифицированную Мегабазу - с учетом дополнений за 2018 год (9,74 миллиона партий), чтоб не терять время на заливку дополнений, можно скачать по ссылке

<https://yadi.sk/d/S7lIdoYf5GWVZ1g>

Особая ценность Мегабазы – в возможности генерирования на ее основе достоверной суперкниги дебютов, имеющей примерно такой вид –



The screenshot shows a chess software interface. On the left is a chessboard with pieces in their starting positions. On the right is a statistics table for a database named 'Megabook.ctg'. The table has columns for move, N (number of games), % (percentage), Av (average), Perf (performance), Fact (fact), Prob (probability), and [%] (percentage in brackets). The table lists various opening moves and their statistics.

	N	%	Av	Perf	Fact	Prob	[%]
Megabook.ctg	9626007	54	585	609			
1.e4!*	4585859	54	1063	1094	0	39.4	70.8
1.d4!?	3233884	55	656	681	0	31.9	0
1.Nf3!?	801615	55	2381	2405	0	11.8	0
1.c4!	708213	54	2378	2399	0	10.4	22.6
1.g3	81733	53	2390	2400	0	2.2	3.6
1.f4	53450	48	2192	2199	0	1.1	1.8
1.Nc3!?	48558	51	2316	2331	0	1.4	0
1.b3!?	45317	50	2308	2307	0	1.1	0
1.b4	29471	48	2137	2148	0	0.7	1.2
1.e3?!	9874	40	2031	2022	0	0	0
1.g4?	7391	43	2223	2184	0	0	0
1.d3!?	5890	46	2074	2075	0	0	0
1.c3?!	4841	47	2067	2081	0	0	0
1.a3?!	4041	48	2135	2139	0	0	0
1.h3	2253	48	2320	2307	0	0	0
1.h4?!	1270	39	1964	1959	0	0	0
1.a4	819	40	1943	1941	0	0	0
1.Nh3?	649	39	1985	1969	0	0	0

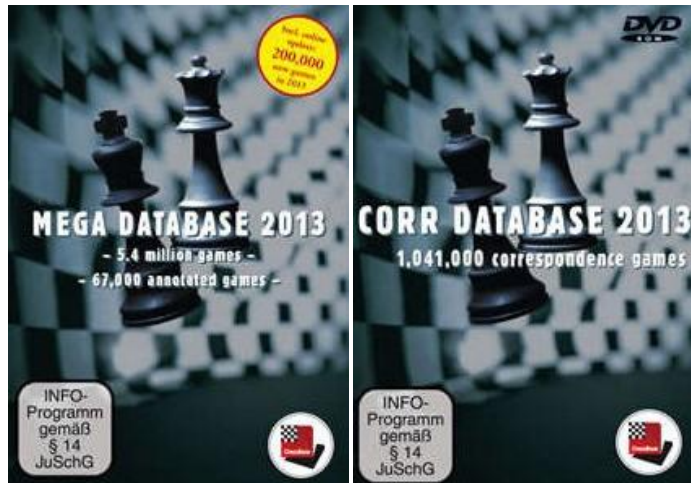
Summary statistics at the bottom right:

- 1-0: 1782700 = 39% N = 4585859 (459)
- 1/2: 1354904 = 30% 1063-3408666
- 0-1: 1448255 = 32% 1066-3409780

ДЛЯ ТЕХ, КТО НЕ В КУРСЕ – ПРОЧИТАЙТЕ РАЗЪЯСНЕНИЕ НИЖЕ –

Дорогие пользователи!

Приобрел в свое время DVD-диски известнейших баз партий, и уже потирал руки от предвкушения пользования ими для успешной заочной игры. Своя супербаза, собираемая и пестуемая около 10 лет, у меня была (7,3 миллиона партий), однако была скомпилирована из десятков источников, а с именами в ней был такой хаос, что имя того же Корчного в ней писалось четырьмя вариантами.



Но...

Начали появляться вопросы, на которые рождались ответы, неудобные для производителя. Что же меня не устроило?

Ответ – не очевидный, но самая главная претензия одна: ОБОИМИ этими базами НЕЛЬЗЯ всерьез пользоваться для подготовки к серьезным турнирам. Чтоб изучить творчество конкретного соперника – это почти пожалуйста (термин «почти» расшифрую ниже), а вот сгенерировать с помощью этих баз (в оболочках Fritz, Shredder) две дебютные книги, оценкам которых можно доверять (а такие суперкниги дебютов по разным дебютам уже висят на торрент-трекерах!) – я бы категорически не рекомендовал. Причина – в обеих базах полным-полно (как оказалось – не менее 35%!) партий крайне слабых любителей, с рейтингом от 400 до 1600 единиц, взятых составителями базы с какого-либо школьного первенства Индонезии, либо, условно говоря, девичьего чемпионата категории до 10 лет не особо шахматного Марокко. Что может нам продемонстрировать это, с позволения сказать, творчество юных и не очень любителей? Ну уж если и не зевки фигур в дебюте, хотя и такое возможно, но в любом случае – не напряженный, мягко скажем, полет дебютной мысли. Конечно, родителям «юных дарований» сам факт наличия творчества их чад в единой мегабазе наверняка греет душу, но нам-то они зачем? ТАКИЕ ПАРТИИ ДЛЯ ПОСТРОЕНИЯ ДЕБЮТНОЙ КНИГИ НЕ НУЖНЫ, ПОТОМУ КАК ОНИ ВСЕРЬЕЗ СМАЖУТ СТАТИСТИКУ ЛЮБОГО ВАРИАНТА – И ТАКОЙ КНИГОЙ БУДЕТ ПОЛЬЗОВАТЬСЯ ПРОСТО ОПАСНО!

Расшифрую, как обещал, смысл и вышесказанного «почти пожалуйста». Я столкнулся с безобразным фактом – фамилии шахматистов в обеих базах этих баз пишутся по-разному! От 2-х – до 5-ти вариантов написаний! И если вы наберете в поисковике, к примеру, «Rich...», вы рискуете никогда не увидеть имя того же игрока в транслитерации «Rych...». Кроме того, и это особенно неприятно для тех, кто платит за эти продукты приличные деньги, в базах полным-полно дублей! Дублей

партий – около 10%, а также – имен! Вот, неверующие могут любоваться или проверить сами:

Bibas	Gratuliano	14
Bibas	Gratuliano J N	12
Bibas	Gratuliano JN	6

Понятно, что это – не три разных человека, а один, поданный под тремя «соусами» написания в базе троекратно. А если, как это в базе сплошь и рядом, он еще в варианте «Bhibas», то вы его и не найдете. Вот и более безобразный образец:

Theting	H	8
Theting	Helge	47
Theting	Helge (NOR)	12
Theting	Helge E	42

Очевидно, что составители баз допустили в этом плане небрежности, могли бы нанять десяток студентов, раскидав между ними буквы латиницы, и под руководством опытного шахматного куратора исправить эти дефекты.

Кроме того, в базе очных партий полным-полно (более 105 тысяч) игр с ...нулевым количеством ходов, где в швейцарках играл «Vue», их, конечно, тоже не мешало бы с базы вовремя убрать. Не верите – в фильтре поиска партий введите число ходов от 0 до 1, при этом цифра 0 мышкой не вводится, а вбивается только вручную.

Также не понравилось то, что для того, чтобы познакомиться в полном объеме с творчеством конкретного шахматиста, надо рыскать по обоим этим базам. Ведь десятки тысяч шахматистов, в том числе ну очень известных, баловались не просто игрой по переписке, но и до сих пор играют заочно на сервере ИКЧФ, интернет-сайтах, и не только блиц. Пример, который напрашивается – Вячеслав Рагозин, хотя есть не столь громкие имена...

Но... Что есть – то есть, однако в результате приходит на ум единственное решение – объединить обе этих супербазы вместе и самостоятельно эти шероховатости устранить. С тем, чтобы потом, «на выходе», получить единую Объединенную Унифицированную Мегабазу, и как результат, ДОСТОВЕРНУЮ дебютную книгу, так и навести вручную (вместо так и не нанятых фирмой-производителем студентов) порядок с именами.

Слил обе базы в одну, на выходе, как и предполагал, получил ожидаемый результат, можно любоваться:

Ragozin	Viacheslav	711
Ragozin	Vyacheslav	37

Понятно, что тот же Рагозин сыграл 711 партий за доской и 37 – по переписке. Однако в ряде случаев, что было уже приятно, имена, не отличавшиеся по транслитерации в обеих базах, в единой - слились в одну.

Но для формирования полноценной единой супербазы, а на ее основе – единой дебютной суперкниги - которые, как представляется, не будут иметь аналогов, я предпринял следующие шаги:

А. Удалены результаты с «шахматистом» «Вуе».

Б. Удалены все партии, завершившиеся вничью до 12-го хода. Отсутствие гроссмейстерских ничьих, может, и исказит на выходе статистику тех или иных шахматистов, однако дебютная книга будет информативнее и достовернее. Ведь даже у талантливых шахматистов бывали дни, когда (в силу турнирного положения или отсутствия настроения) в шахматы им играть не хотелось – они и не играли. Более 2-х десятков подобных эпизодов – в карьере того же Корчного.

В. Удалены все партии шахматистов с рейтингом ниже 2100 единиц, и здесь пришлось попотеть, так как приходилось учитывать тот факт, что система ЭЛО введена сравнительно недавно, в 1979 году, так что ни Петросян, ни Таль, творившие в 70-е годы и не имевшие рейтинга, не «пострадали». Да, можно было бы включить в фильтре поиска партии с рейтингом от 2100 до 2900 и сделать выборку, но в нее не вошли бы ни Таль, ни Спасский. Такая выборка не нужна. А если же включить в фильтре поиска партии с рейтингом от нуля до 2099 (для их отсека), то в него укладываются абсолютно все гроссы и чемпионы мира, игравшие в безрейтинговые эпохи. Поэтому здесь пришлось работать вручную, путем визуального просмотра. Работа заняла 15 месяцев, в результате оставлены все партии безрейтинговых шахматистов, также игравших по переписке или под никами на серверах. Ведь отсутствие в подобных случаях рейтинга вовсе не означает, что игрок играет слабо, или что под ником не скрывается мастер или гросс. Здесь, согласитесь, лучше все же недобдеть, чем перебдеть - и отсесть десятки тысяч вполне приличных для дебютной теории партий. Оставлены все партии ПО ПЕРЕПИСКЕ, а те, что игрались в интернете, отсекались по низшему рейтингу хотя бы одного партнера 1200. Почему такой низкий уровень? С учетом того факта, что ВСЕ в интернете играют при помощи комп-программ (интернет-блиц отсекал весь), и того, что на ряде серверов, того же LSS, свои внутренние системы рейтингов, ниже официальных.

Г. Отдельный вопрос – компьютерное творчество. При решении этого вопроса я исходил из той логичной аксиомы, что ЛЮБОЙ движок, за небольшими исключениями, играет практически уж никак не менее силы обозначенного мною рейтинга 2100 – и поэтому компьютерные партии в мою единую базу мною включены. Причем ВСЕ блиц-партии (тестовые в том числе) движков мною ИСКЛЮЧЕНЫ, оставлены лишь те, что шли с совокупным контролем от 25 минут на партию и выше. Таковых оказалось немного - 2,6 миллиона из примерно 6 миллионов, болтающихся на различных сайтах для свободного скачивания.

Д. Партии вслепую (тот же Амбер-турнир) и блиц-партии «белковых» шахматистов в единой базе оставлены. Понятно, что в миттельшпиле в подобных случаях ведущие шахматисты могут не блистать, но уж дебют, в любом случае, поставят солидно и в полную силу.

При идентификации и унификации имен шахматистов учитывались рейтинги, годы игры, более поздний (последний) вариант написания имени, особенно пришлось попотеть над ирландскими и азиатскими именами. При этом удалялись точки, скобки, приставки «dr», «prof», у славянских шахматистов удалены отчества. Например, «Foigel Igor» объединялся со строкой «Foygel Igor», «Francev GN» - со строкой «Frantsev Gennady». В ходе работы унифицированная база дополнена мною сотнями партий, найденных в своей базе, имен шахматистов, игравших в 18-начале 20-го века - от 15 до 50 партий для каждого. Объединены 35000 дублей по именам,

удалены партии 27500 шахматистов, играющих на любительском уровне. Вот примеры тех, что удалялись:

Thogersen Rasmus 4

Белые	Эло Б	Черные	Эло Ч	Результ...	Год
Thogersen,R		Marshall,M	1901	0-1	2010
Kyrkjebo,E	1922	Thogersen,R		1-0	2010
Thogersen,R		Rolvaag,B	1919	0-1	2010
Holm,K	2086	Thogersen,R		1-0	2010

У шахматиста в базе 4 партии, все проиграны в одну калитку примерно 3-м перворазрядникам и 1-му КМСу.

Tivane Graca da Felecida.. 3

Белые	Эло Б	Черные	Эло Ч	Результ...	Год
Tivane,G		Latreche,S	1985	0-1	2011
Tivane,G		Pasku,R	1912	0-1	2010
Tivane,G		De Seroux,C	2020	0-1	2010

У шахматистки в базе 3 партии, все проиграны в одну калитку примерно 2-м перворазрядникам и 1-му КМСу.

Feltrin Kris 3

Белые	Эло Б	Черные	Эло Ч	Результ...	Год
Fantini,M	1997	Feltrin,K		1-0	2010
Brank,B	1901	Feltrin,K		1-0	2010
Lakovic,I	1934	Feltrin,K		1-0	2010

У шахматиста в базе 3 партии, все проиграны в одну калитку примерно 3-м перворазрядникам.

Но при этом если я видел, что шахматист без рейтинга или с рейтингом, играющий в нынешнюю, уже рейтинговую эпоху, имел победы над соперниками с рейтингом от 2100 и выше, то его партии я оставлял, как, например, в этом случае:

Titov Sergo 7

Белые	Эло Б	Черные	Эло Ч	Результ...	Год
Ibragimov,A	2094	Titov,S	1906	S-S	2009
Medvedev,M	2177	Titov,S	1906	1-0	2009
Shebchenko,V	1977	Titov,S	1906	S-S	2009
Titov,S	1906	Domnin,V	2016	S-S	2009
Titov,S	1906	Kukushkin,S	2015	1-0	2009
Titov,S	1906	Lazarenko,V	2090	0-1	2009
Titov,S	1906	Medvedev,V	2160	S-S	2009

У шахматиста в базе 7 партий, несмотря на невысокий личный рейтинг, есть ничьи и победы над КМСами