

## Модель оценки сложности физической схемы реляционной базы данных

*А.А. Рыбанов, О.В. Свиридова, Н.Н. Короткова, Д.Н. Лясин,*

*О.Ф. Абрамова*

*Волжский политехнический институт (филиал) Волгоградского государственного  
технического университета, г. Волжский Волгоградской обл.*

**Аннотация:** В работе рассмотрены области применения количественных метрик баз данных. Контроль качества схемы базы данных невозможен без числовых показателей. Метрики базы данных можно вычислять автоматизированно, что гарантирует точность и повторяемость таких измерений. В статье предлагается модель для оценки сложности физической схемы реляционной базы данных, основанная на метриках таблицы базы данных. Методом Саати получены коэффициенты модели, учитывающие степень влияния каждой метрики на сложность базы данных. На примере реальных проектов баз данных проведены количественные измерения их метрик в MySQL. Описаны результаты сравнительного анализа, полученных по модели, оценок сложности различных проектов физических схем баз данных и количества информации, содержащейся в sql-скриптах этих базы данных.

**Ключевые слова:** оценка сложности, база данных, физическая схема, модель, SQL, энтропия

### Введение

Метрики являются полезными механизмами для улучшения качества программных продуктов [1]. К сожалению, почти все показатели ориентированы на оценку качества программного кода и практически не применимы к базам данных (БД). Традиционными индикаторами "качества" реляционных баз данных являются оценка нормализации схемы базы данных и верификация содержащейся в базе данных информации [2, 3].

В тоже время, контроль качества схемы базы данных невозможен без числовых показателей. При отсутствии количественных измерений трудно принимать какие-либо проектные решения [4]. Знание размера, сложности и качества унаследованных БД необходимо для их преобразования и повторного использования. Для построения объективного представления о базе данных необходимо использовать связанный набор метрик, которые будут отражать целостное представление о качестве базы данных [5, 6].

## Постановка проблемы

Метрики базы данных можно вычислять автоматизированно, что гарантирует точность и повторяемость таких измерений, а также позволяет строить на их основе различные аналитические отчеты [7].

В MySQL с помощью запросов к информационной базе данных INFORMATION\_SCHEMA, которая хранит информацию относительно всех других баз данных, могут быть получены следующие исходные количественные метрики для  $i$ -ой таблицы физической схемы базы данных: количество атрибутов в таблице ( $m_{i,1}$ ); количество ключей, включая уникальные индексы ( $m_{i,2}$ ); количество внешних ключей ( $m_{i,3}$ ); количество неуникальных индексов ( $m_{i,4}$ ); количество различных типов данных ( $m_{i,5}$ ); количество атрибутов, входящих в состав первичного ключа ( $m_{i,6}$ ); количество атрибутов, входящих в состав уникальных индексов ( $m_{i,7}$ ); количество атрибутов, входящих в состав неуникальных индексов ( $m_{i,8}$ ); количество атрибутов, входящих в состав внешних ключей ( $m_{i,9}$ ); количество таблиц-родителей ( $m_{i,10}$ ); количество таблиц-потомков ( $m_{i,11}$ ).

Данные метрики могут быть использованы для оценки сложности физической схемы базы данных.

Рассмотрим уже существующий алгоритм оценки сложности базы данных [8] в котором для каждой таблицы БД вычисляется вес сложности  $W_i$  по формуле:  $W_i = m_{i,1} + m_{i,2} + m_{i,3} + m_{i,4}$ . Сложность физической схемы базы данных  $S$  вычисляется как сумма весов сложности её таблиц:  $S = \sum W_i$ .

В алгоритме учитываются не все метрические характеристики, а также не учитывается степень влияния каждого параметра на полученный результат.

В работе [9] предлагается модель с большим количеством метрик, но остается нерешенной задача получения коэффициентов влияния этих метрик

---

на сложность базы данных. Актуальной является задача построения модели оценки сложности базы данных с учетом весовых коэффициентов метрик.

### Модель оценки сложности физической схемы базы данных

Предлагается следующая модель для оценки сложности физической схемв базы данных:

$$C = \sum W_i,$$

$$\text{где } W_i = \alpha_1 m_{i.1} + \alpha_2 m_{i.2} + \alpha_3 m_{i.3} + \alpha_4 m_{i.4} + \alpha_5 m_{i.5}.$$

Коэффициенты  $\alpha_k$  определим с помощью процедуры Саати [10]. Матрица парных сравнений метрик оценки сложности таблиц БД, с оценками согласованности  $ИС=0.1259$  и  $ОС=0.1124$ , приведена в таблице №1.

Таблица № 1

Матрица парных сравнений метрик сложности таблиц БД

Метрики таблицы БД	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	Вектор приоритетов	$\alpha_k$
$m_1$	1	1/5	1/7	5	3	0.844	0.105
$m_2$	5	1	1/3	9	7	2.537	0.315
$m_3$	7	3	1	9	5	3.936	0.489
$m_4$	1/5	1/9	1/9	1	1/3	0.242	0.030
$m_5$	1/3	1/7	1/5	3	1	0.491	0.061
Сумма	13.533	4.454	1.787	27.000	16.333	8.050	1

На основании оценки сложности физической схемы БД можно судить о трудоемкости проектных работ, выполненных разработчиком БД. Например, сложность физической схемы БД employees в 1.577 раз больше, чем сложность физической схемы БД music (рис. 1). На основании этого можно косвенно оценить, что трудоемкость работ по созданию БД employees примерно в 1.577 раз больше, чем для БД music.

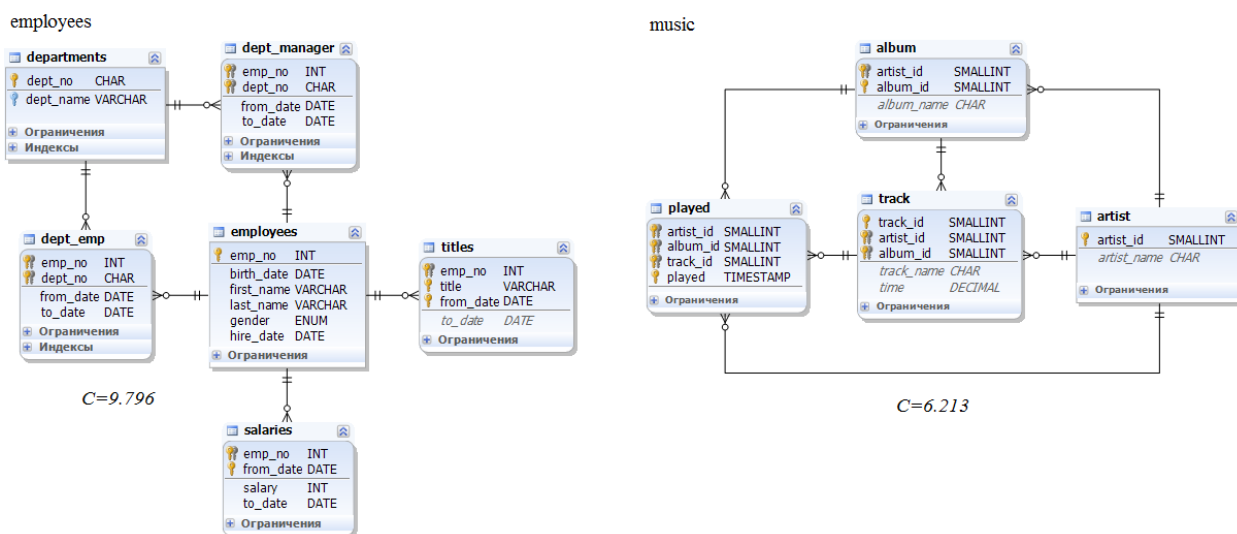


Рис. 1. – Физические схемы БД employees и БД music

В таблице №2 приведены результаты расчета сложности для физических схем различных проектов БД по предлагаемой модели.

Таблица №2

Результаты расчета сложности различных физических схем БД

Наименование проекта БД	Аконим	Сложность БД	Нормализованное значение сложности БД
flight	FL	4.894	0.0262
world	WO	5.113	0.0274
music	MU	6.213	0.0333
employees	EM	9.796	0.0525
university	UN	8.589	0.0600
classicmodels	CL	13.843	0.0741
retailer	RE	14.920	0.0787
chinook	CH	16.900	0.0905
contracts	CO	18.734	0.1003
northwind	NO	23.151	0.1240
sakila	SA	35.246	0.1887
catalog	CA	29.576	0.1584

Сравнительный анализ полученных результатов

Выполним сопоставление, полученных по модели, оценок сложности физических схем баз данных и количества информации, содержащейся в sql-скриптах этих базы данных. Для проведения количественного анализа sql-

скрипта физической схемы базы данных воспользуемся понятием информационной энтропии. Информационная энтропия для случайных и независимых значений  $x_i$  с  $M$  возможными состояниями рассчитывается по

$$\text{формуле: } H(x) = -\sum_{i=1}^M p_i \cdot \log_2 p_i .$$

Для исследования sql-скрипта физической схемы базы данных с помощью понятия информационной энтропии необходимо определить роль и диапазон значений величины  $x$ . Sql-скрипт базы данных на самом низком уровне своей реализации представляет собой поток байтов, в котором  $x$  может принимать значения в диапазоне  $[0;255]$ , следовательно, в роли вероятностей  $p_i$  будет выступать частота повторов конкретного байта.

Информационная энтропия – это статистический параметр, который показывает вероятность встречаемости определённых байтов в файле. На рис. 2 приведена гистограмма распределения повторов одинаковых байтов в sql-скрипте физической схемы базы данных, которая позволяет визуально оценить информационную энтропию.

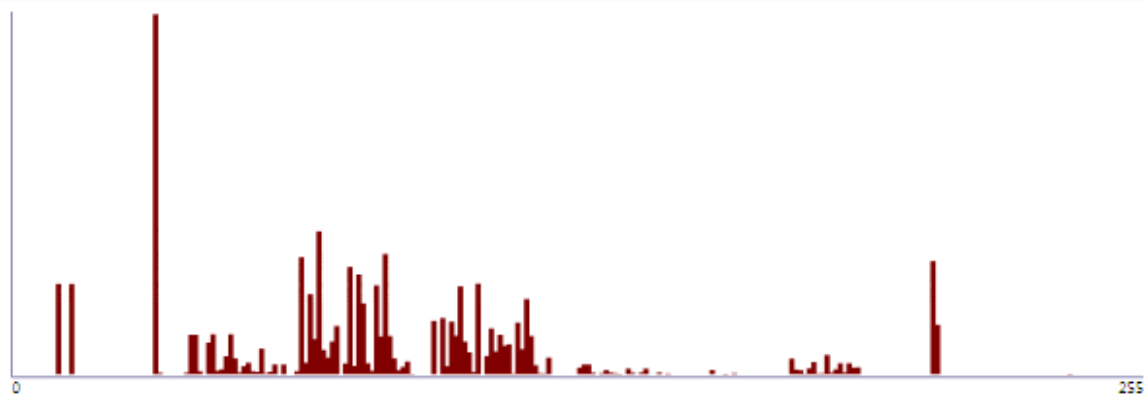


Рис. 2. – Гистограмма sql-скрипта физической схемы БД Chinook: ось  $X$  - значение байта, ось  $Y$  - количество байтов с заданным значением

Количество информации в сообщении, содержащем  $n$  символов  $I_n$ , по

$$\text{Шеннону равно: } I_n = -n \sum_{i=1}^M p_i \cdot \log_2 p_i .$$

В таблице №3 приведены результаты расчета энтропии и количества информации для sql-скриптов физических схем различных проектов БД.

Таблица №3

Результаты расчета информационной энтропии для различных проектов БД

Наименование проекта БД	Акроним	Энтропия sql-скрипта, бит	Количество информации в sql-скрипте, бит	Нормализованное значение количества информации
flight	FL	5.8862	15733.8102	0.0351
world	WO	5.8200	16703.3192	0.0373
music	MU	5.7430	18061.8459	0.0403
employees	EM	5.6903	19751.0573	0.0440
university	UN	5.7335	22051.0135	0.0492
classicmodels	CL	5.6597	32916.7644	0.0734
retailer	RE	5.6084	35714.4131	0.0797
chinook	CH	5.6008	47589.8962	0.1061
contracts	CO	5.8200	52246.0959	0.1165
northwind	NO	5.5462	53099.1014	0.1184
sakila	SA	5.4819	66430.1994	0.1482
catalog	CA	5.6548	68084.1489	0.1518

Сопоставление оценок сложности физической схемы БД и количества информации, содержащейся в sql-скрипте БД приведено на рис. 3.

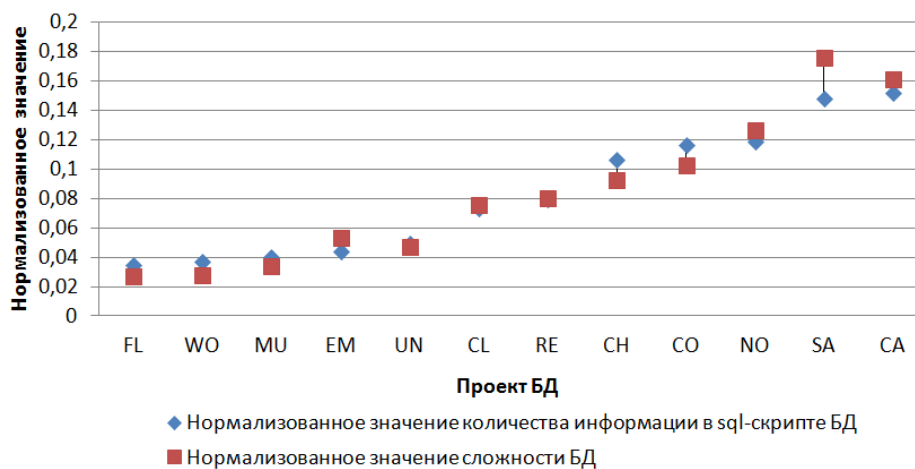


Рис. 3. – Сравнительный анализ нормализованных значений сложности БД и, соответствующих им значений, количества информации sql-скрипта БД

Значение коэффициента корреляции Пирсона  $r=0.9662$  говорит о наличии сильной связи между нормализованными значениями оценок сложности БД и количеством информации в sql-скрипте БД.

### **Заключение**

Наиболее точную оценку сложности реляционной базы данных можно получить в результате анализа её физической схемы. Предложенная модель оценки сложности физической схемы БД может быть использована для объективной косвенной оценки трудоемкости выполненных разработчиком базы данных проектных работ. Рассмотренный подход к автоматизированному получению количественных метрик, описывающих физическую схему реляционной БД, может быть взят за основу веб-ориентированной информационной системы количественной оценки физических схем реляционных базы данных.

### **Литература**

1. Звездин С.В. Проблемы измерения качества программного кода // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2010. № 2 (178). С. 62-66.
2. Галушка В.В., Молчанов А.А., Фатхи В.А. Применение многослойных радиально-базисных нейронных сетей для верификации реляционных баз данных // Инженерный вестник Дона. 2012. №1. URL: [ivdon.ru/ru/magazine/archive/n1y2012/686](http://ivdon.ru/ru/magazine/archive/n1y2012/686).
3. Наумов А.А., Айдинян А.Р. Надежность программного обеспечения и методы ее повышения // Инженерный вестник Дона. 2018. №2. URL: [ivdon.ru/ru/magazine/archive/N2y2018/4946](http://ivdon.ru/ru/magazine/archive/N2y2018/4946)



4. Черняев А.О., Рыбанов А.А. Разработка и исследование алгоритмов автоматизированного проектирования логических схем реляционных баз данных // В мире научных открытий. 2010. № 4-11 (10). С. 128-129.

5. Piattini M., Calero C., Genero M. Table oriented metrics for relational databases. Software Quality Journal. 2001, №9 (2), pp. 79-97.

6. Genero M., Piattini M., Calero C. Measures to get better quality databases. 2<sup>nd</sup> International Conference on Enterprise Information Systems - ICEIS 2000, Stafford, pp. 49-55.

7. Pavlic M., Kaluza M., Vrcek N. Database complexity measuring method. Proceedings of the ISRM 2002 Conference, Las. Vegas, NV, USA, 2002, pp. 577-583.

8. Рыбанов А.А. Оценка сложности физической схемы реляционной базы данных // Современная техника и технологии. 2014. №9. URL: [technology.snauka.ru/2014/09/4330](http://technology.snauka.ru/2014/09/4330).

9. Вершинникова Л.А. Алгоритм для оценки сложности реляционной базы данных // Актуальные проблемы авиации и космонавтики. 2017. Т. 2. № 13. С. 96-97.

10. Rybanov A.A., Makushkina L.A. Technology of an aprioristic objective assessment of distance course themes complexity based on Saati's algorithm // Journal of Engineering Science and Technology Review. 2016. Vol. 9. № 1. pp. 81-89.

### References

1. Zvezdin S.V. Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Komp'yuternye tekhnologii, upravlenie, radioelektronika. 2010. № 2 (178). pp. 62-66.

2. Galushka V.V., Molchanov A.A., Fathi V.A. Inzhenernyj vestnik Dona (Rus). 2012. №1. URL: [ivdon.ru/ru/magazine/archive/n1y2012/686](http://ivdon.ru/ru/magazine/archive/n1y2012/686).





3. Naumov A.A., Ajdinyan A.R. Inzhenernyj vestnik Dona (Rus), 2018, №2. URL: [ivdon.ru/ru/magazine/archive/N2y2018/4946](http://ivdon.ru/ru/magazine/archive/N2y2018/4946).
4. Chernyaev A.O., Rybanov A.A. V mire nauchnykh otkrytiy. 2010. № 4-11 (10). pp. 128-129.
5. Piattini M., Calero C., Genero M. Quality Journal. 2001, №9 (2), pp. 79-97.
6. Genero M., Piattini M., Calero C. 2<sup>nd</sup> International Conference on Enterprise Information Systems - ICEIS 2000, Stafford, pp. 49-55.
7. Pavlic M., Kaluza M., Vrcek N. Proceedings of the ISRM 2002 Conference, Las. Vegas, NV, USA, 2002, pp. 577-583.
8. Rybanov A.A. Sovremennaya tekhnika i tekhnologii, 2014, №9. URL: [technology.snauka.ru/2014/09/4330](http://technology.snauka.ru/2014/09/4330).
9. Vershinnikova L.A. Aktual'nye problemy aviatsii i kosmonavtiki. 2017. Vol. 2. № 13. pp. 96-97.
10. Rybanov A.A., Makushkina L.A. Journal of Engineering Science and Technology Review. 2016. Vol. 9. № 1. pp. 81-89.