

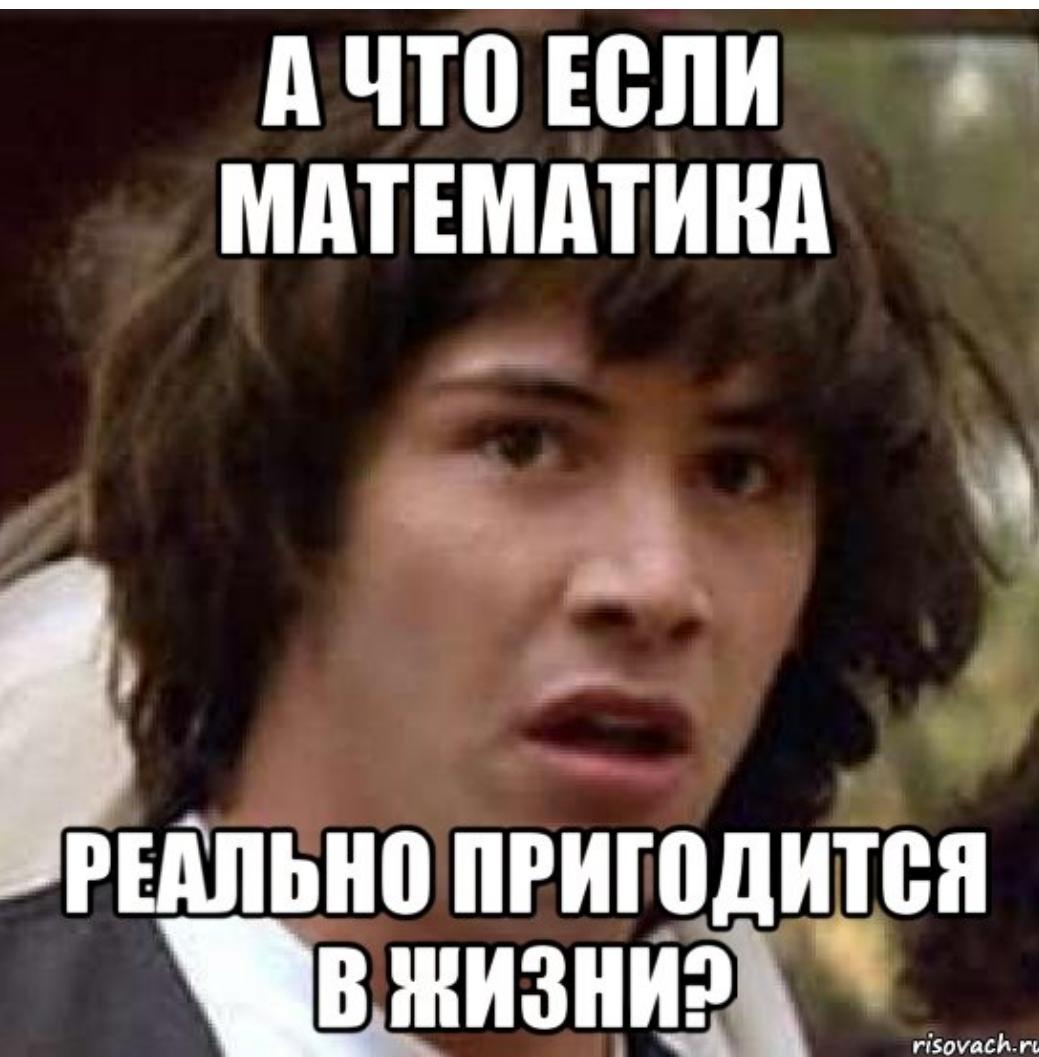
Яндекс



Красивое индексирование в ClickHouse

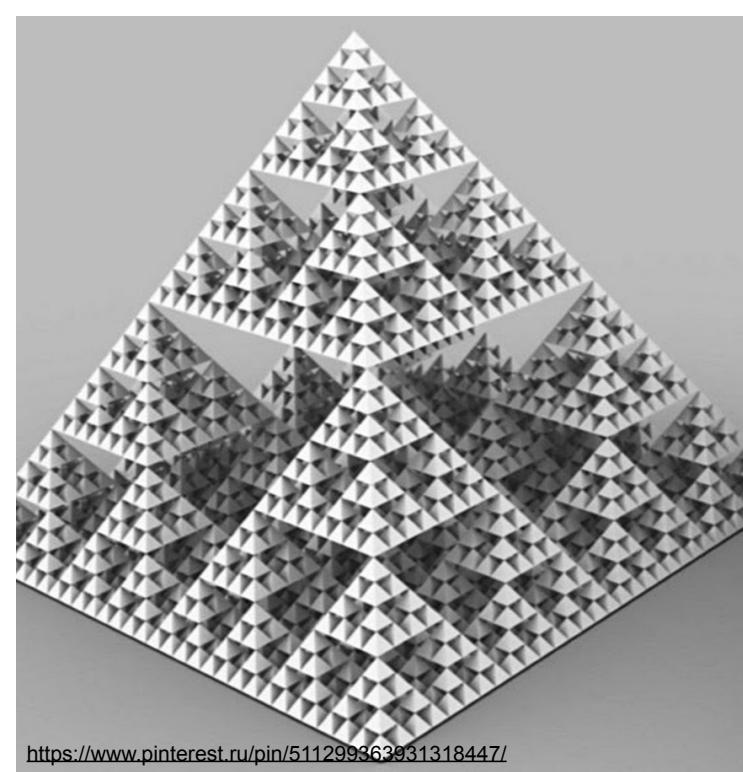
Ольга Хвостикова, разработчик

Красота математики

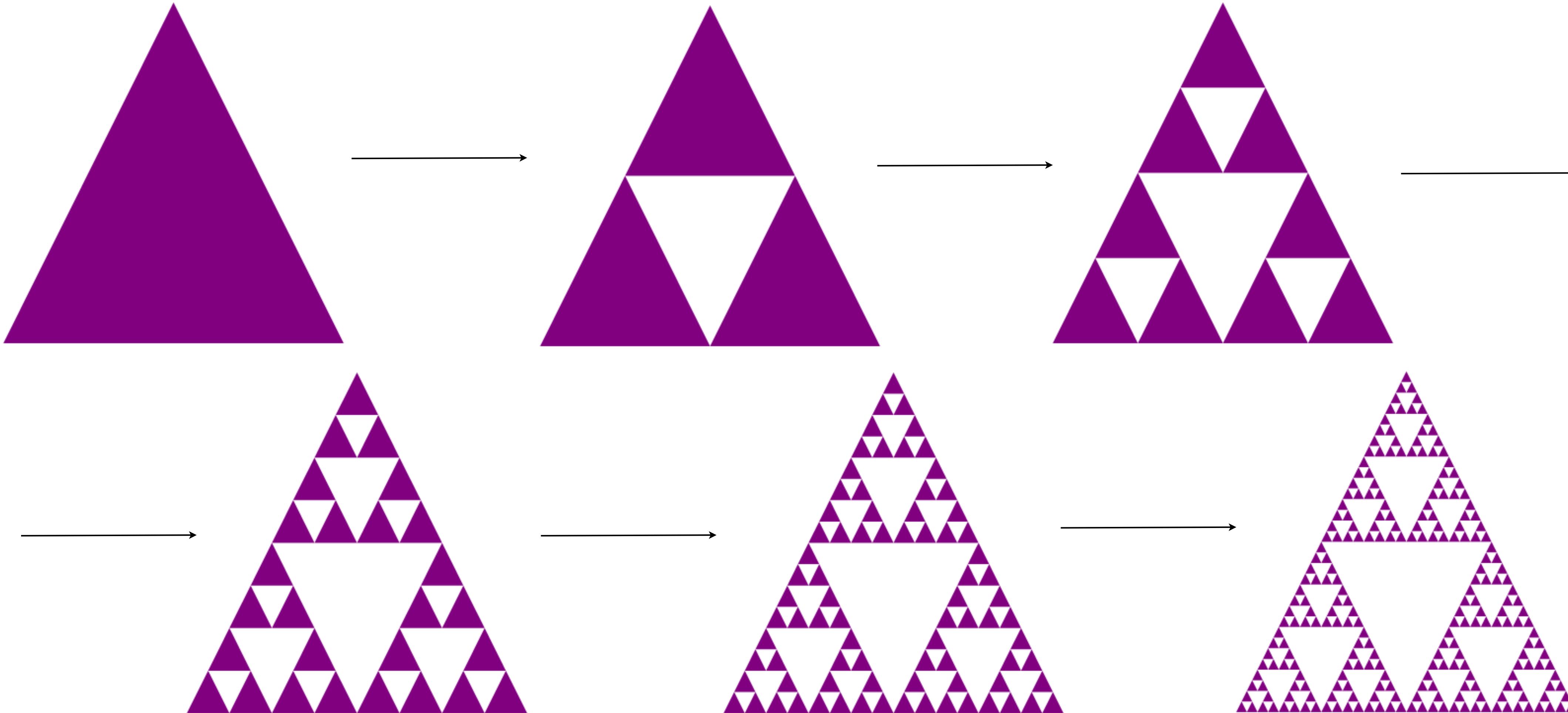


Красота математики

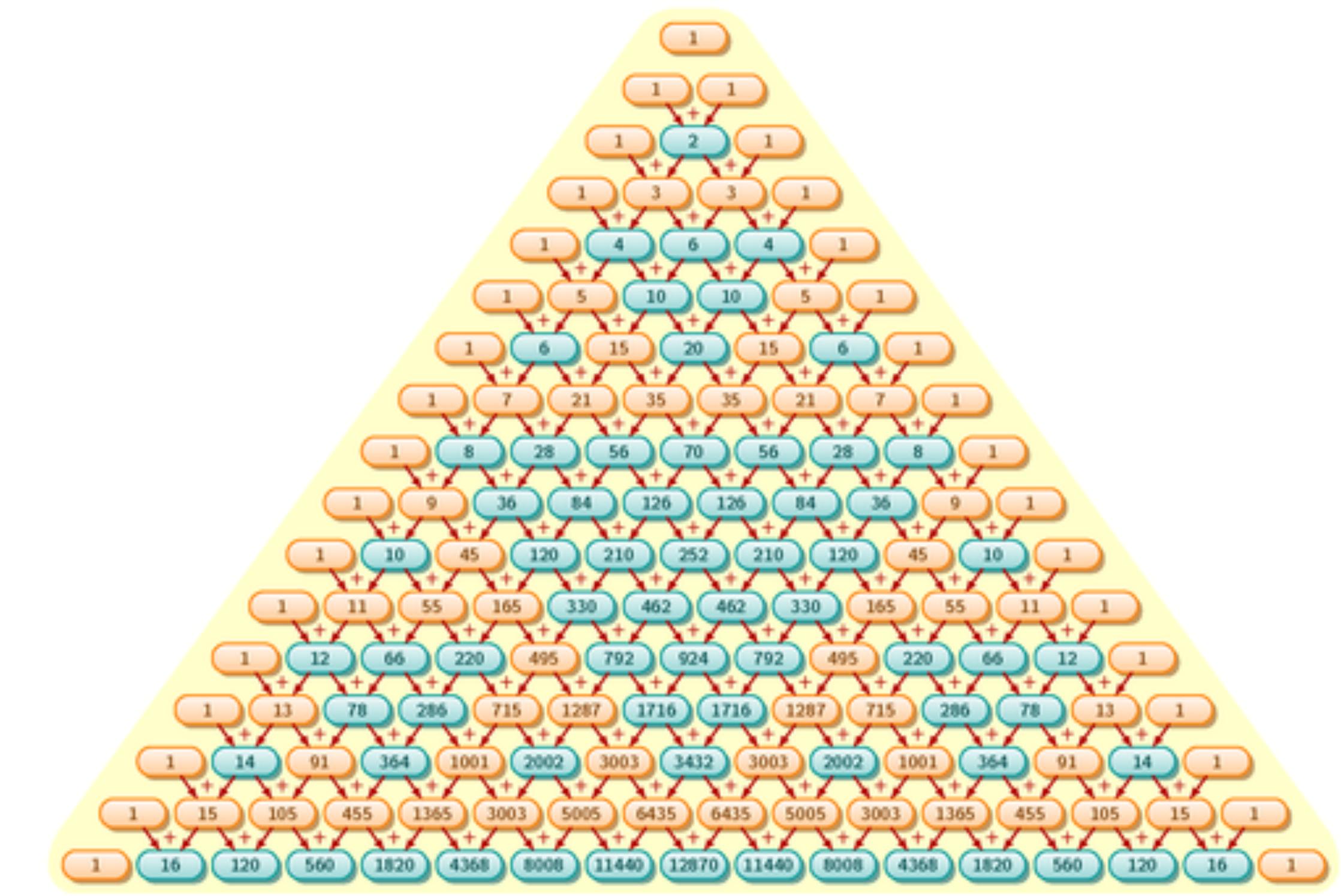
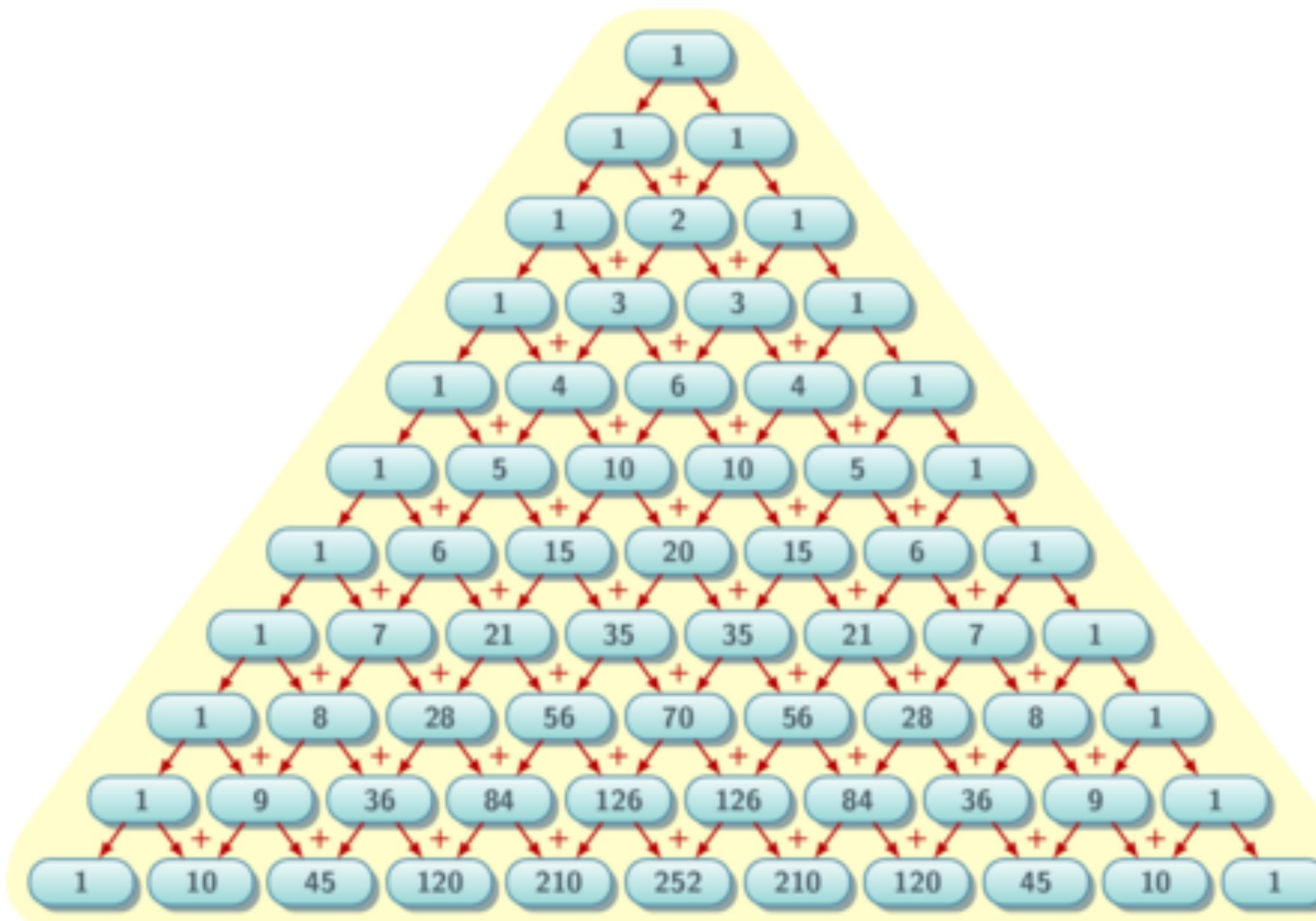
- › Фрактал — фигура, обладающая свойством самоподобия
- › Рекурсия — см. рекурсия



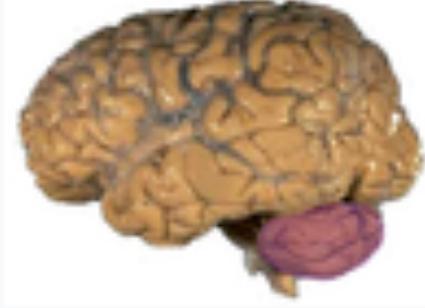
We need to go deeper



Ещё несколько слов о красоте

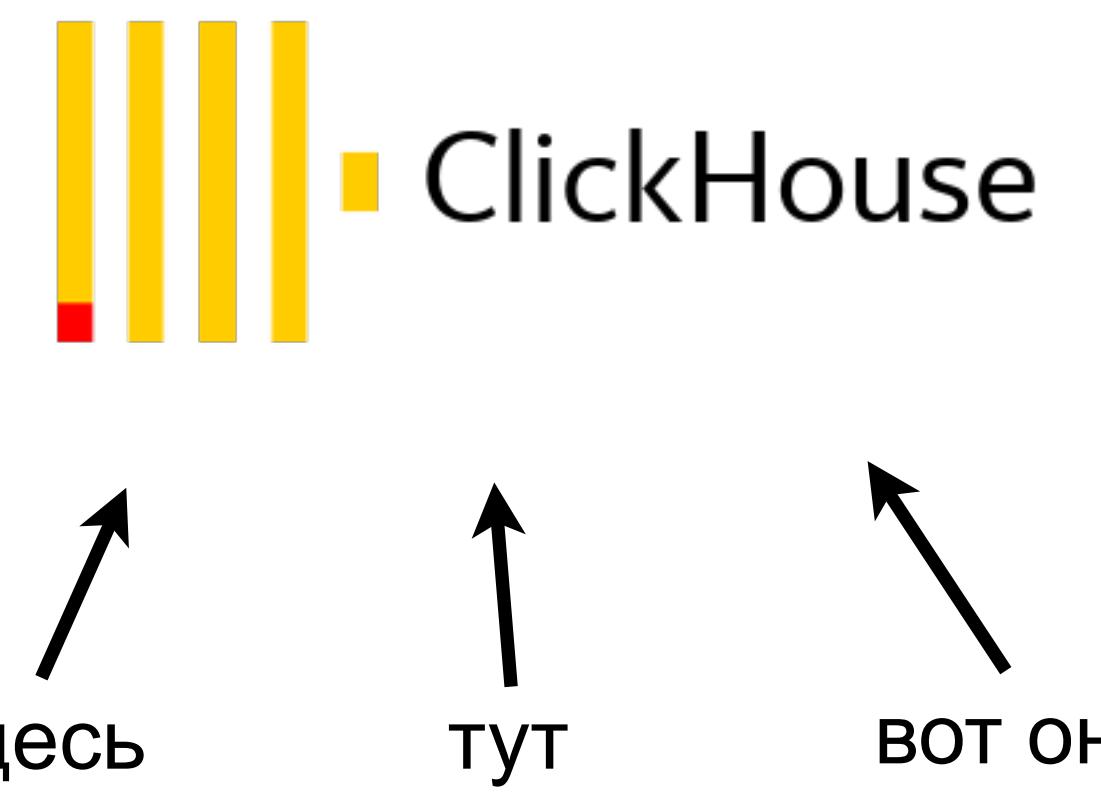


Хаусдорфова размерность

~2.7	The surface of Broccoli	
2.79	Surface of human brain	
~2.8	Cauliflower	

https://en.wikipedia.org/wiki/List_of_fractals_by_Hausdorff_dimension

А где же ClickHouse?



Задача

Хотим хранить многомерные данные.

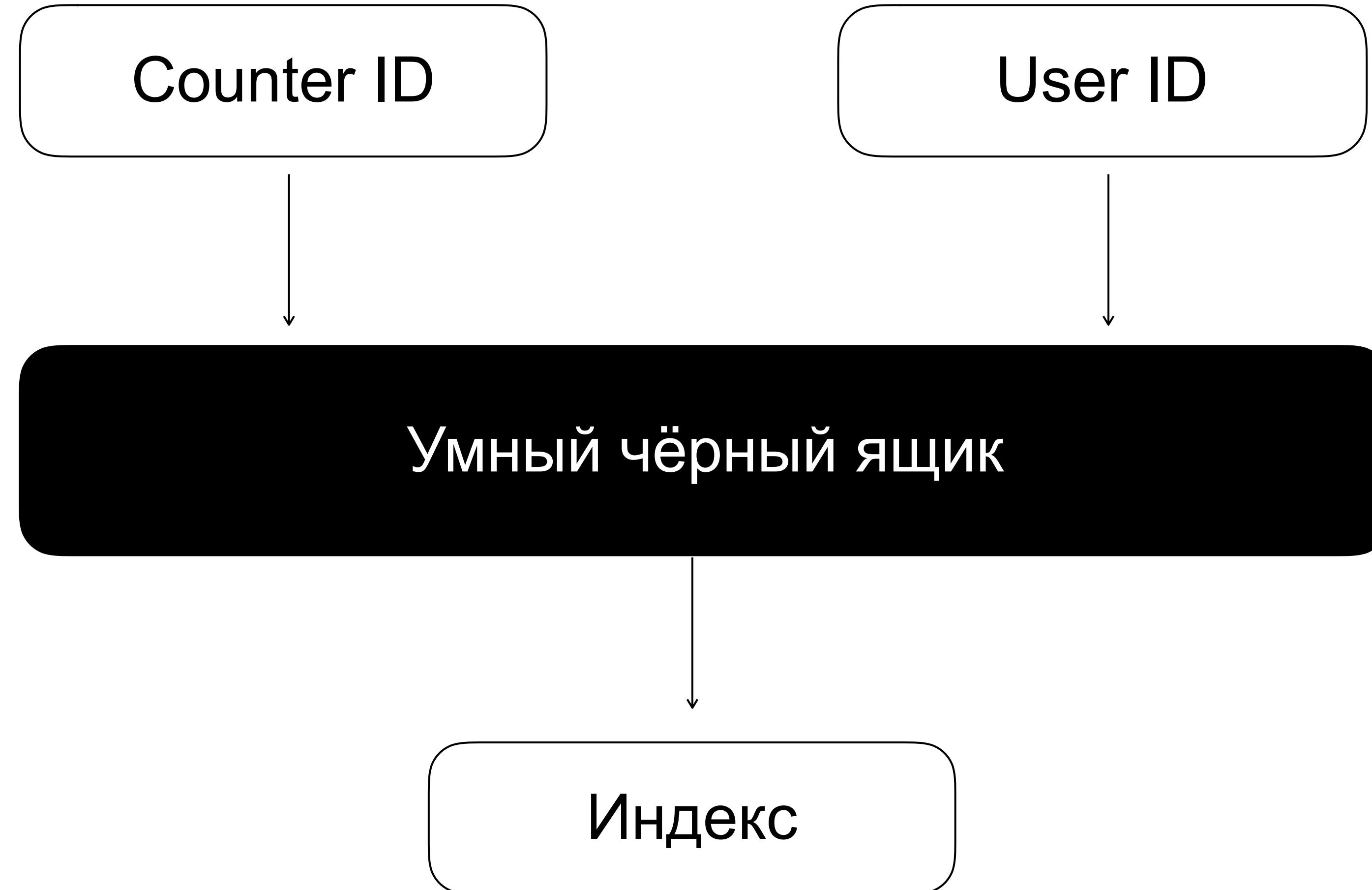
Но без выбора одного единственного первичного ключа, чтобы при сортировке сохранялась локальность по нескольким столбцам.

Пример 1. Red, Green, Blue.

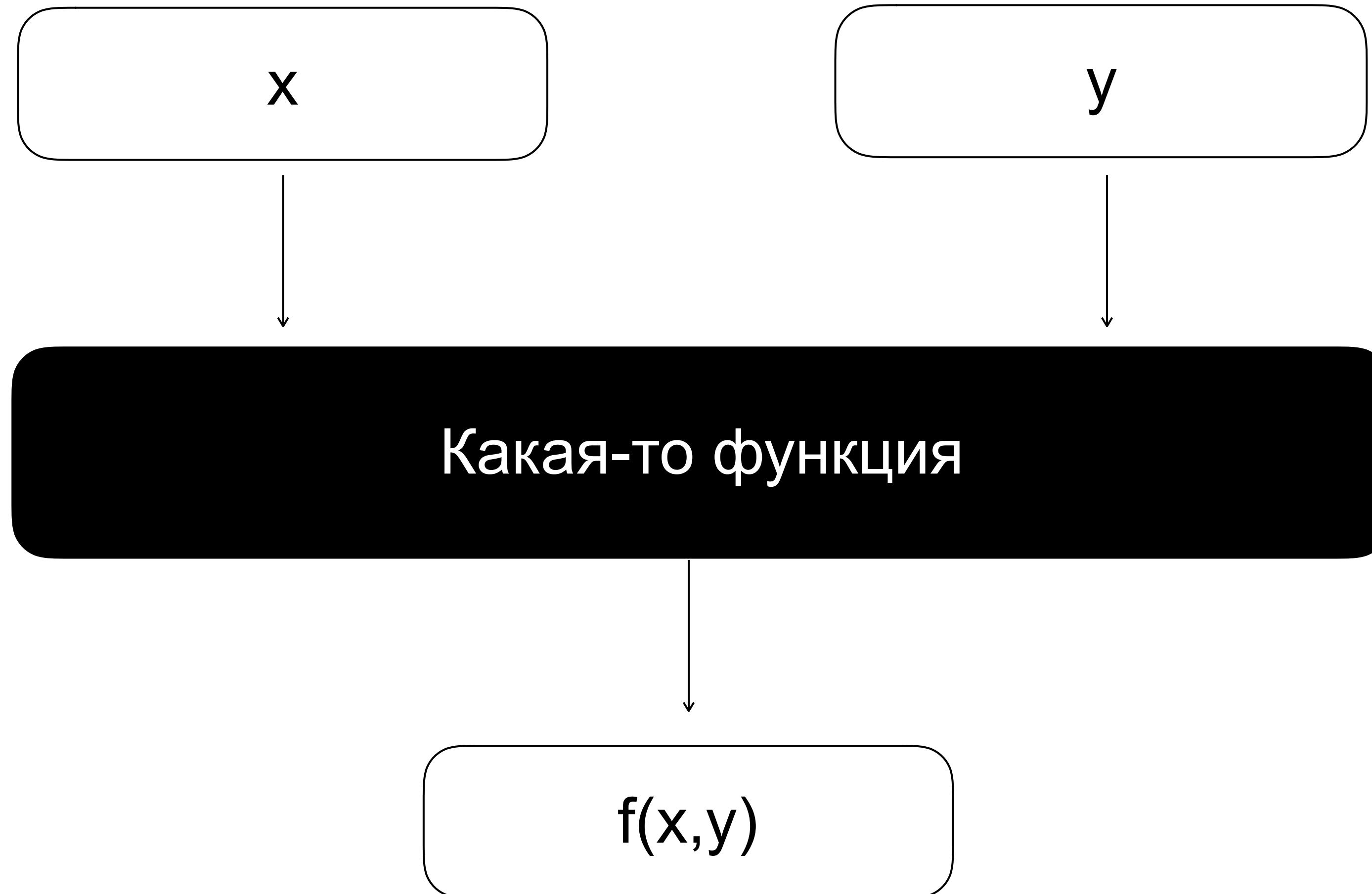
Пример 2. Широта и долгота.

Пример 3. Counter ID и User ID.

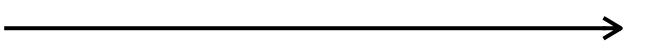
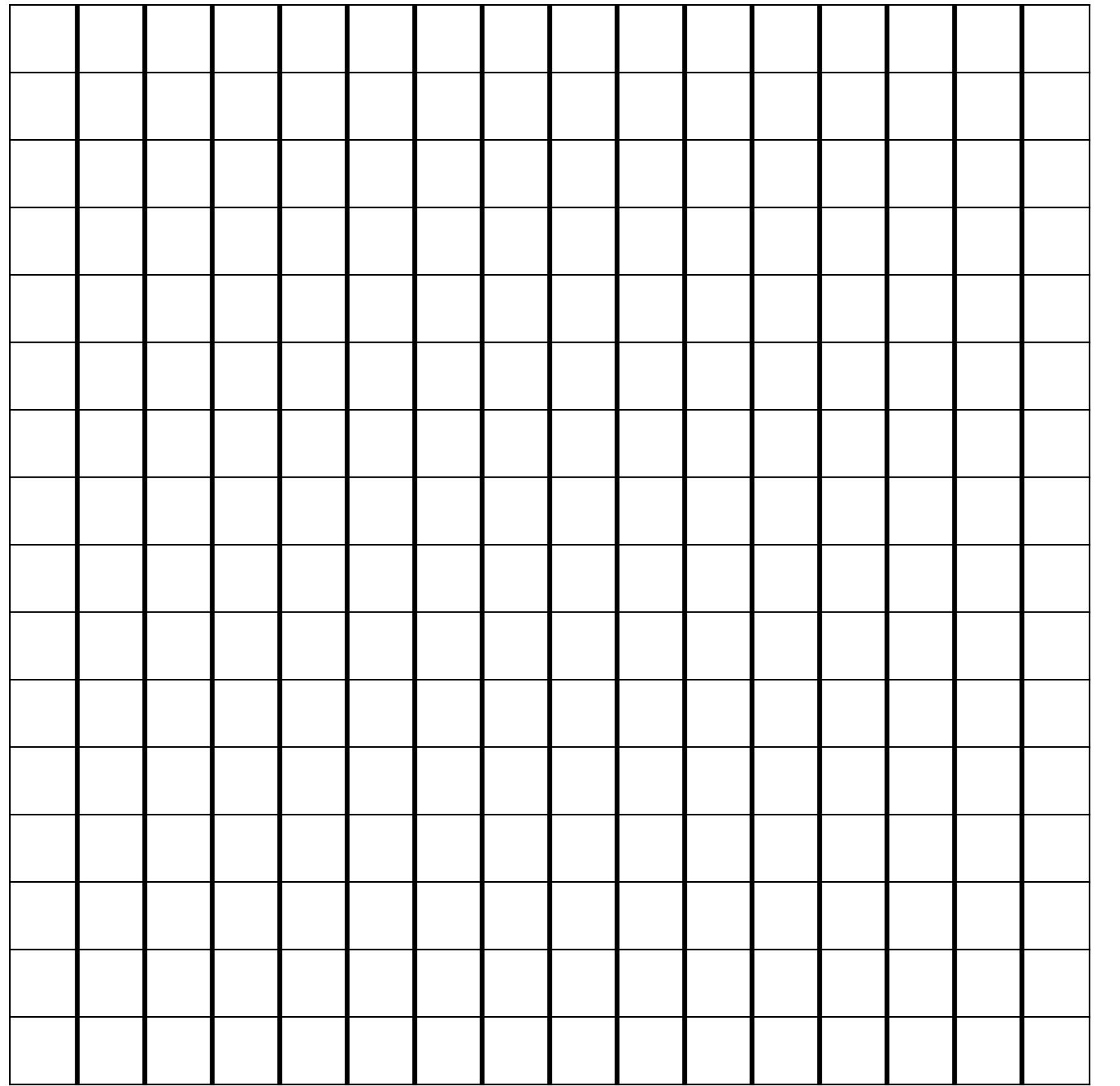
Хотим



Хотим (по-математически)

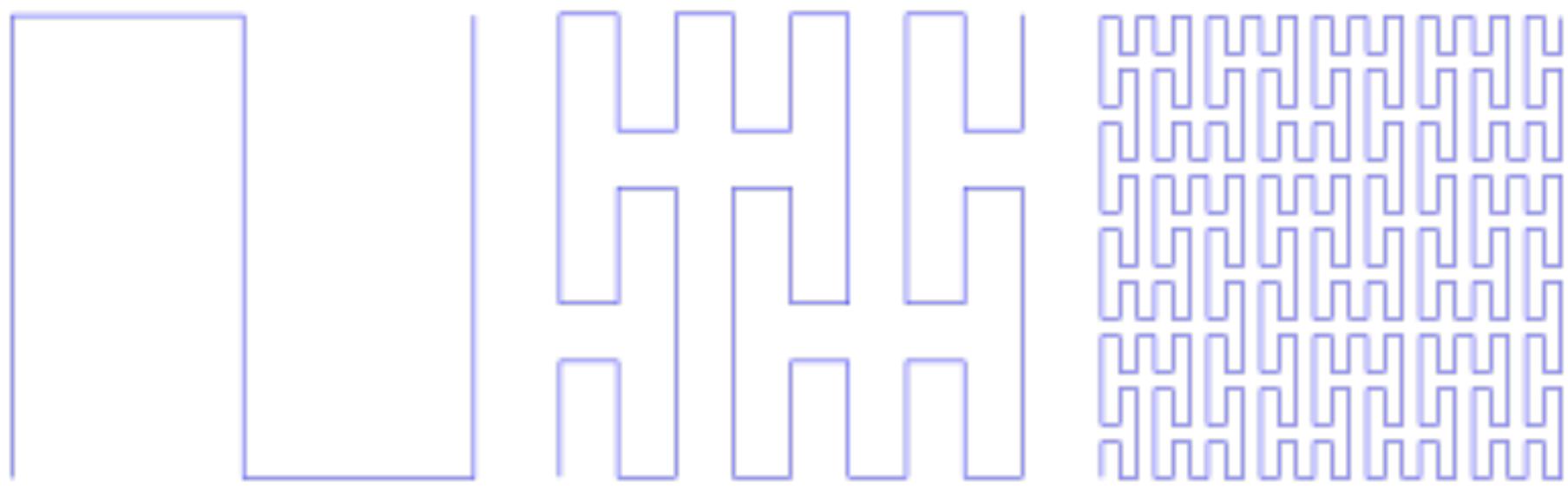


Другими словами



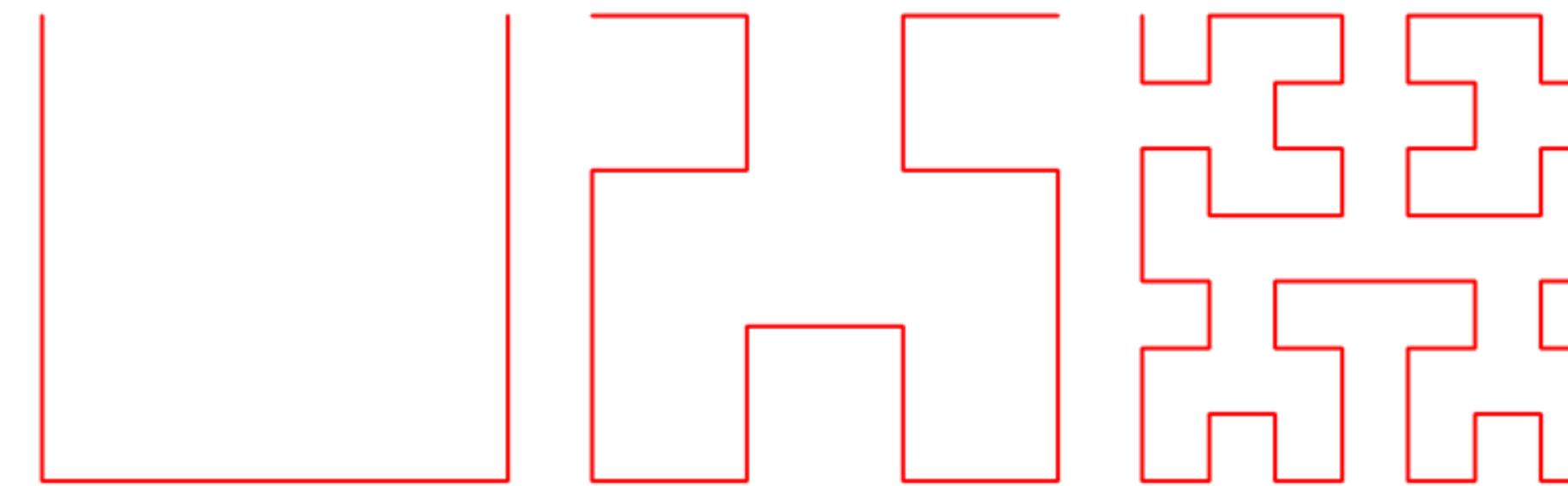
Space filling curves

Кривая Пеано

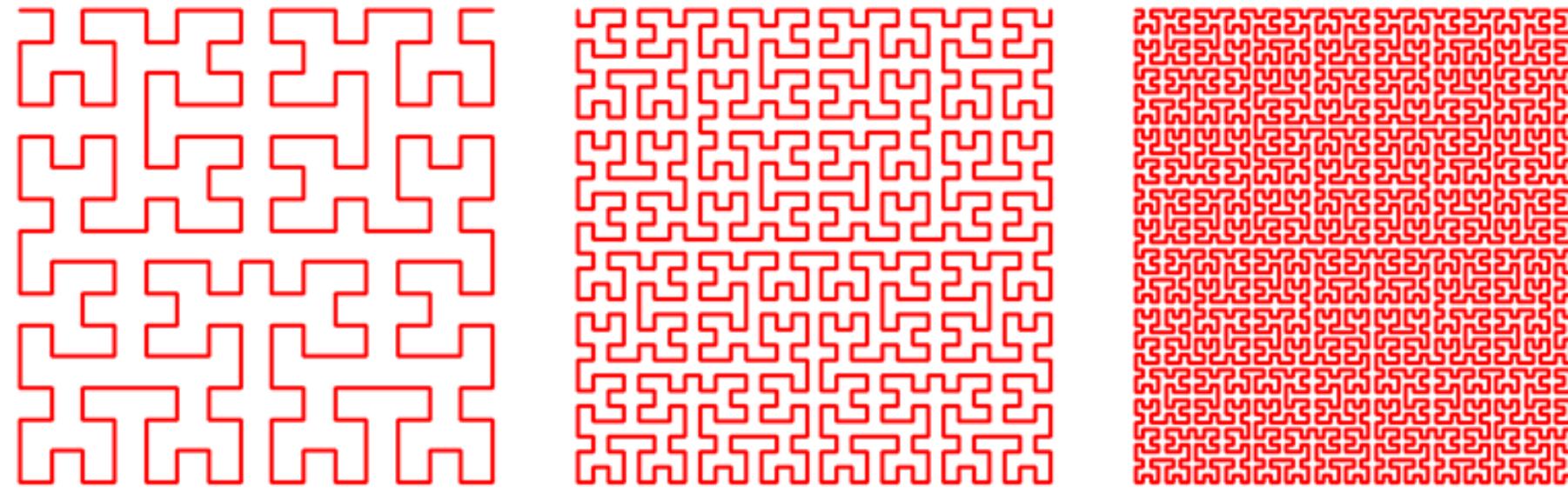


<https://gitak.ru/fraktaly/geometricheskie-fraktaly/830-n-fraktal.html>

Кривая Гильберта

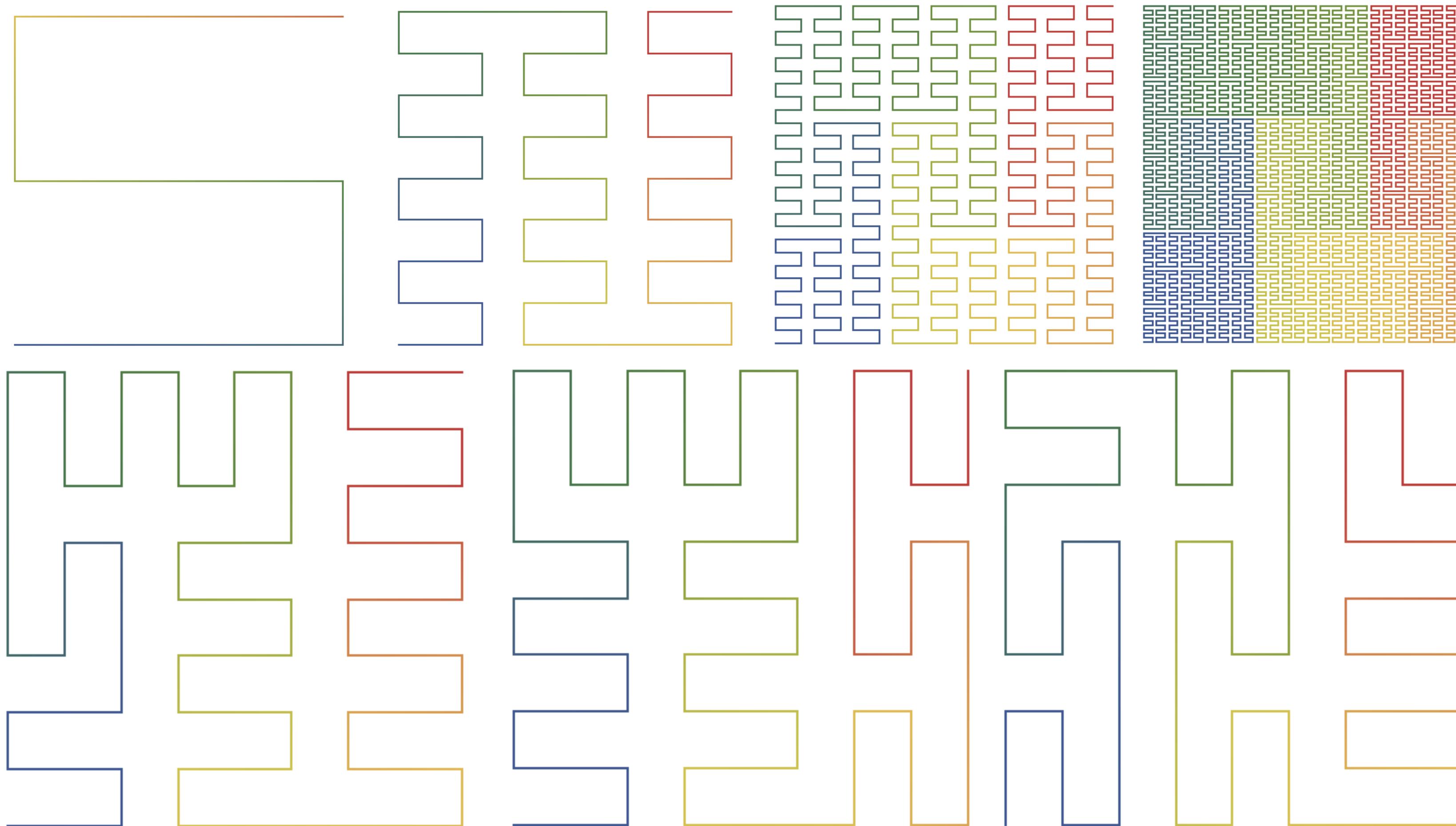


http://hijos.ru/wp-content/uploads/2014/03/hilbert_curve8.png



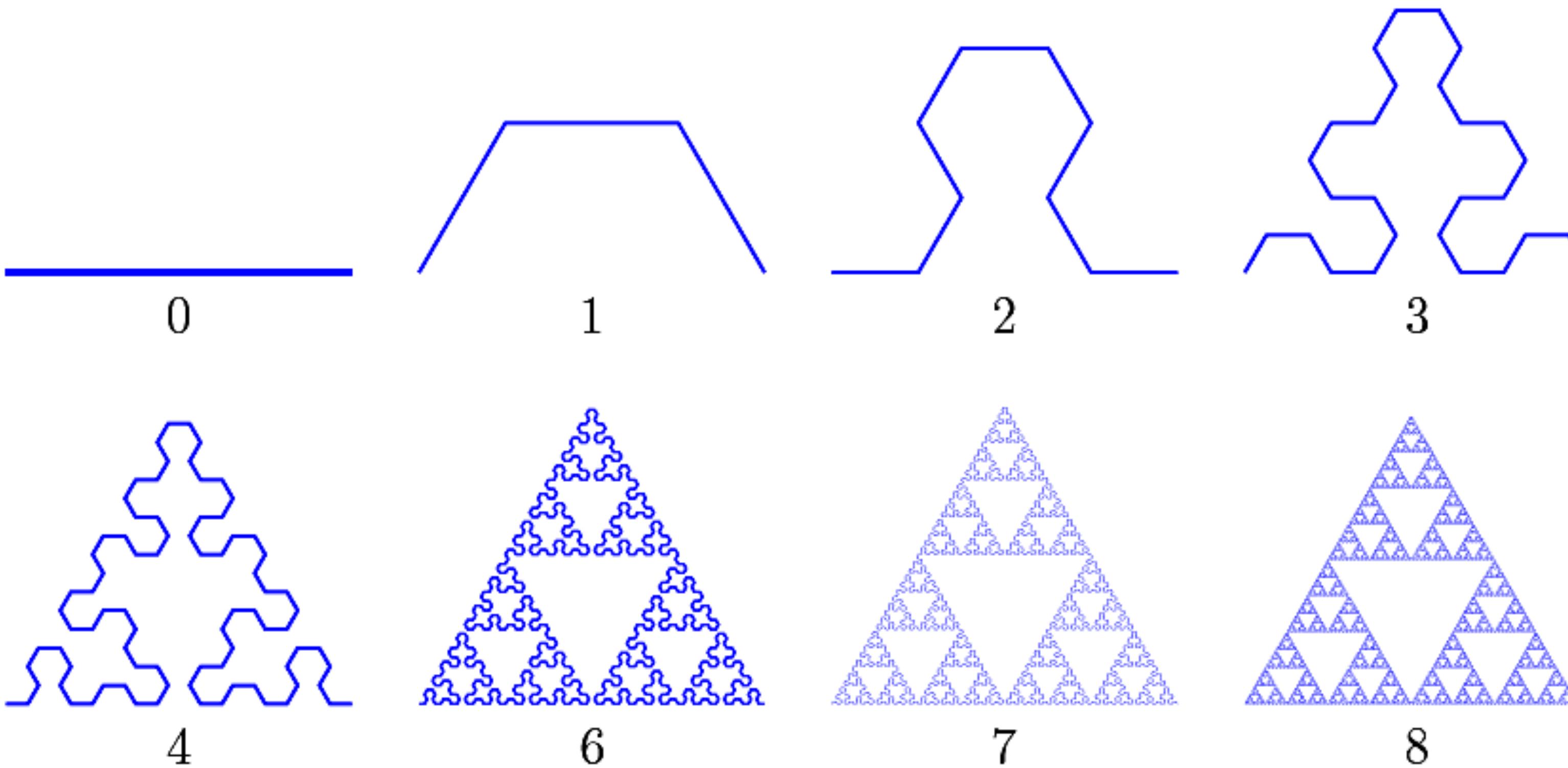
Space filling curves

Кривые Пеано



Space filling curves

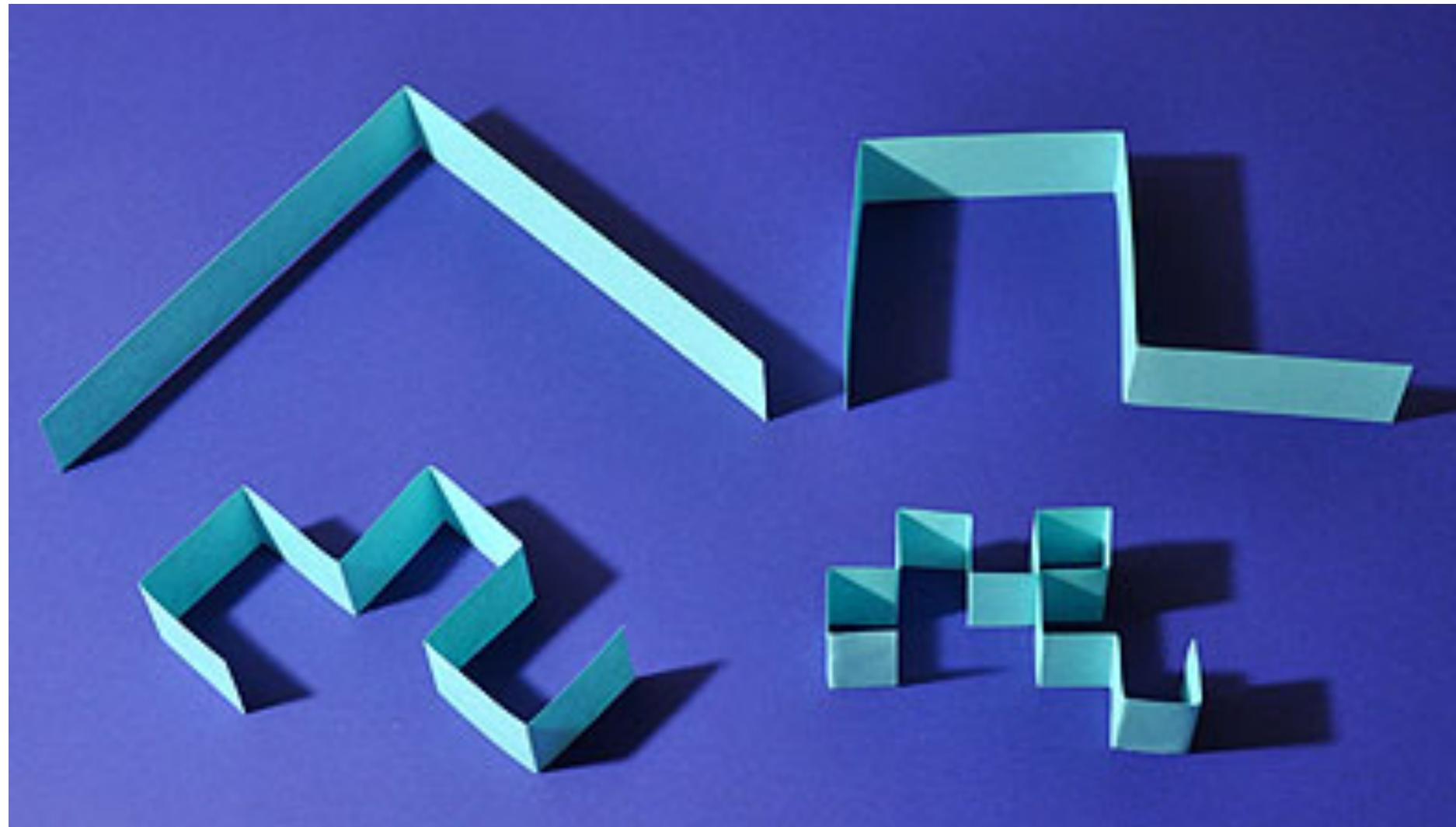
Треугольник Серпинского тоже замещается кривой



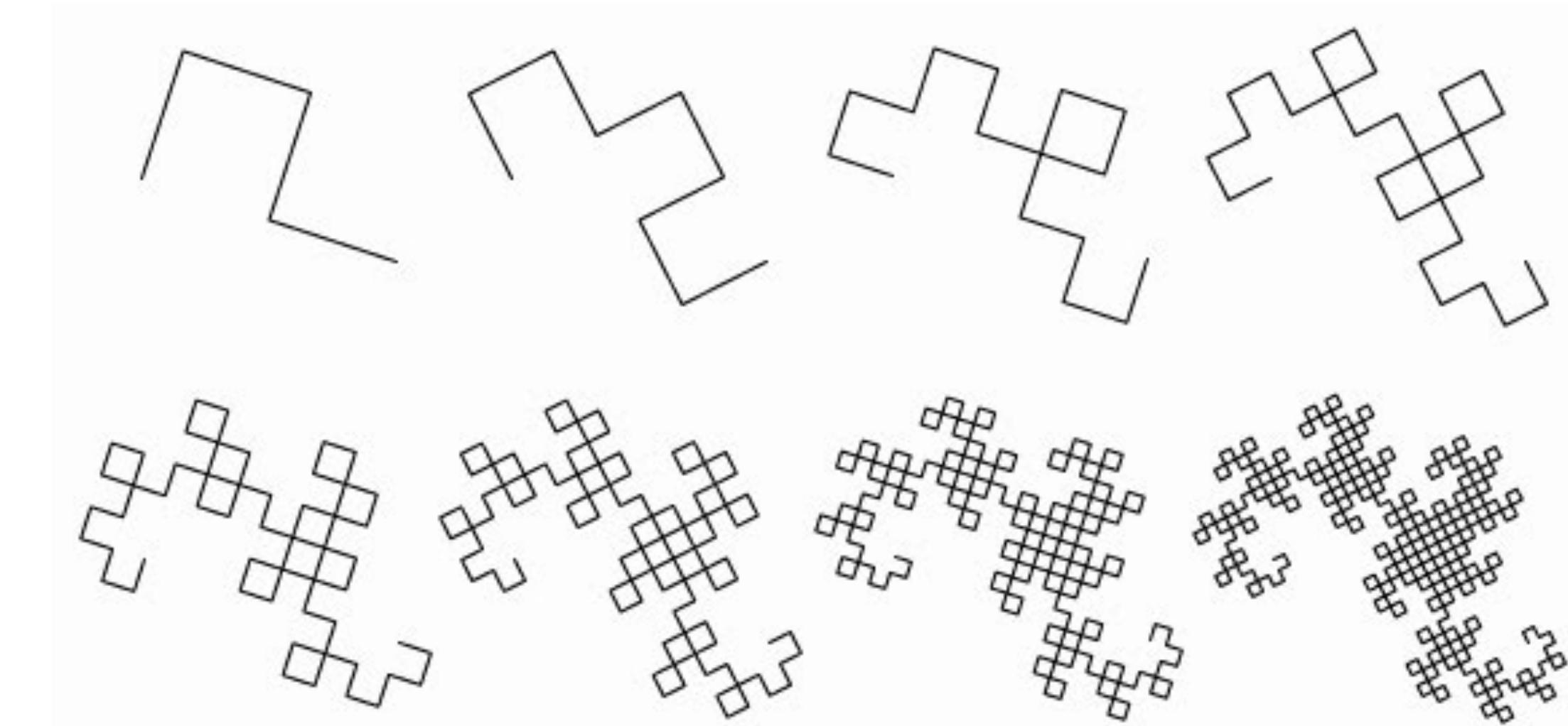
<https://elementy.ru/posters/fractals/Sierpinski>

Space filling curves

Dragon curve

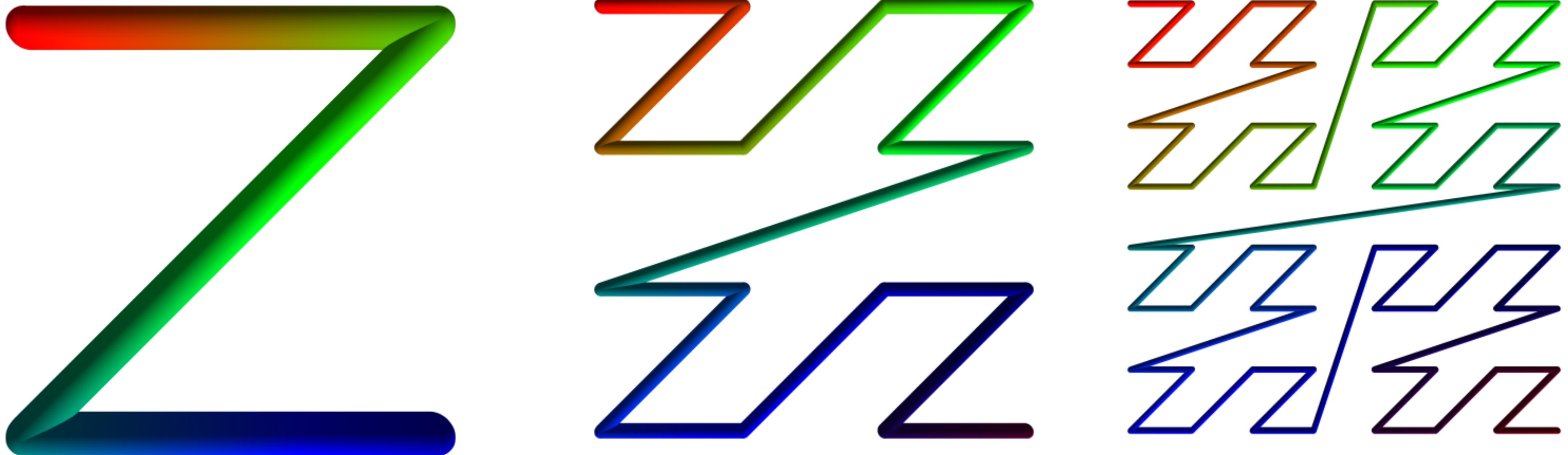


<https://www.cutoutfoldup.com/216-dragon-curve.php>

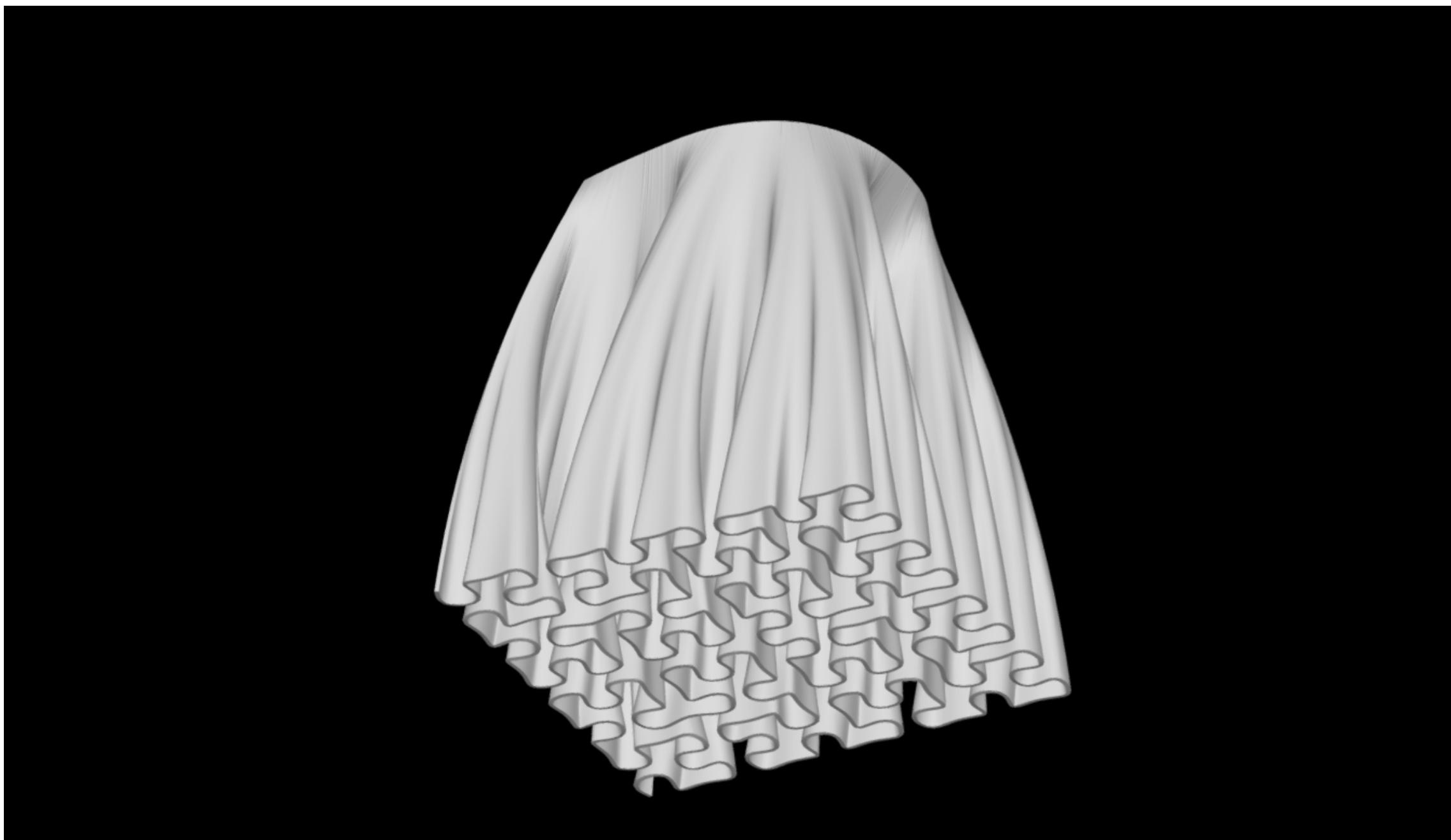


<https://www.physicsforums.com/threads/dragon-curve-fractal-using-golden-ratio.653704/>

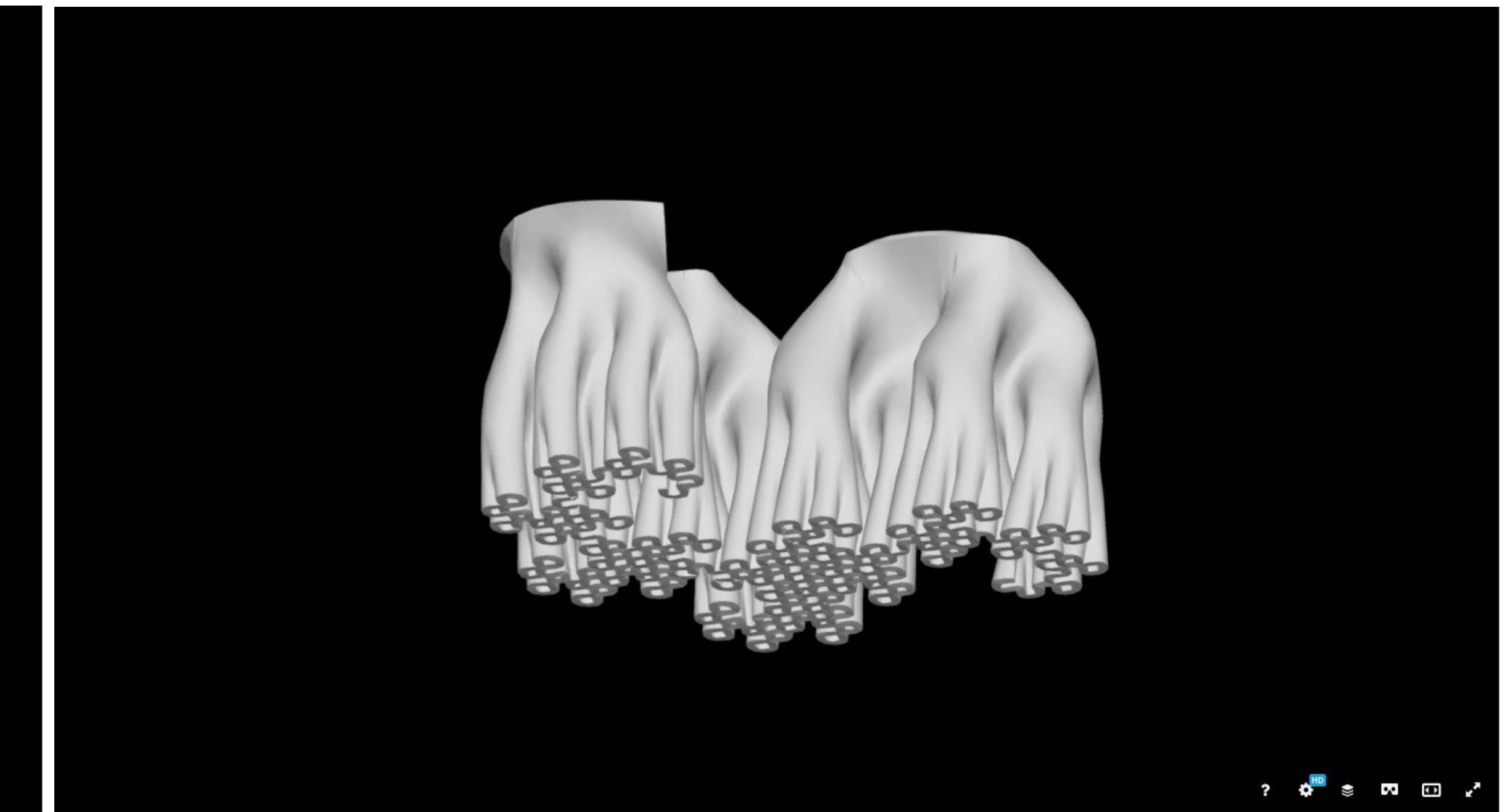
Z-order curve



Многомерность

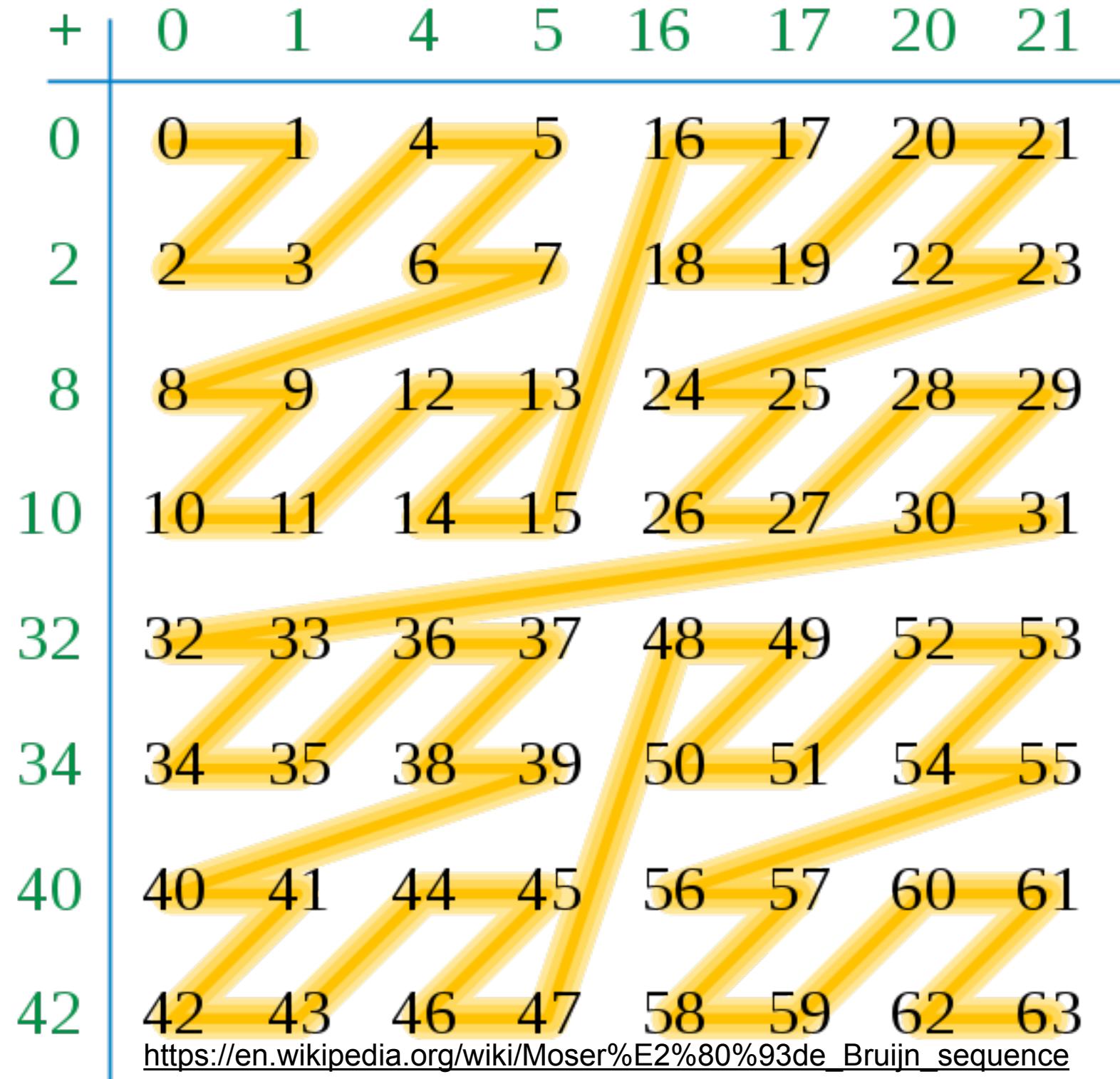


<https://sketchfab.com/3d-models/the-developing-hilbert-curve-hilberts-curtain-8af51753e6874858ab3160d45c0928a8>



<https://www.physicsforums.com/threads/dragon-curve-fractal-using-golden-ratio.653704/>

Z-order curve



i: 1 0 0 1 0

j: 1 1 0 0 1

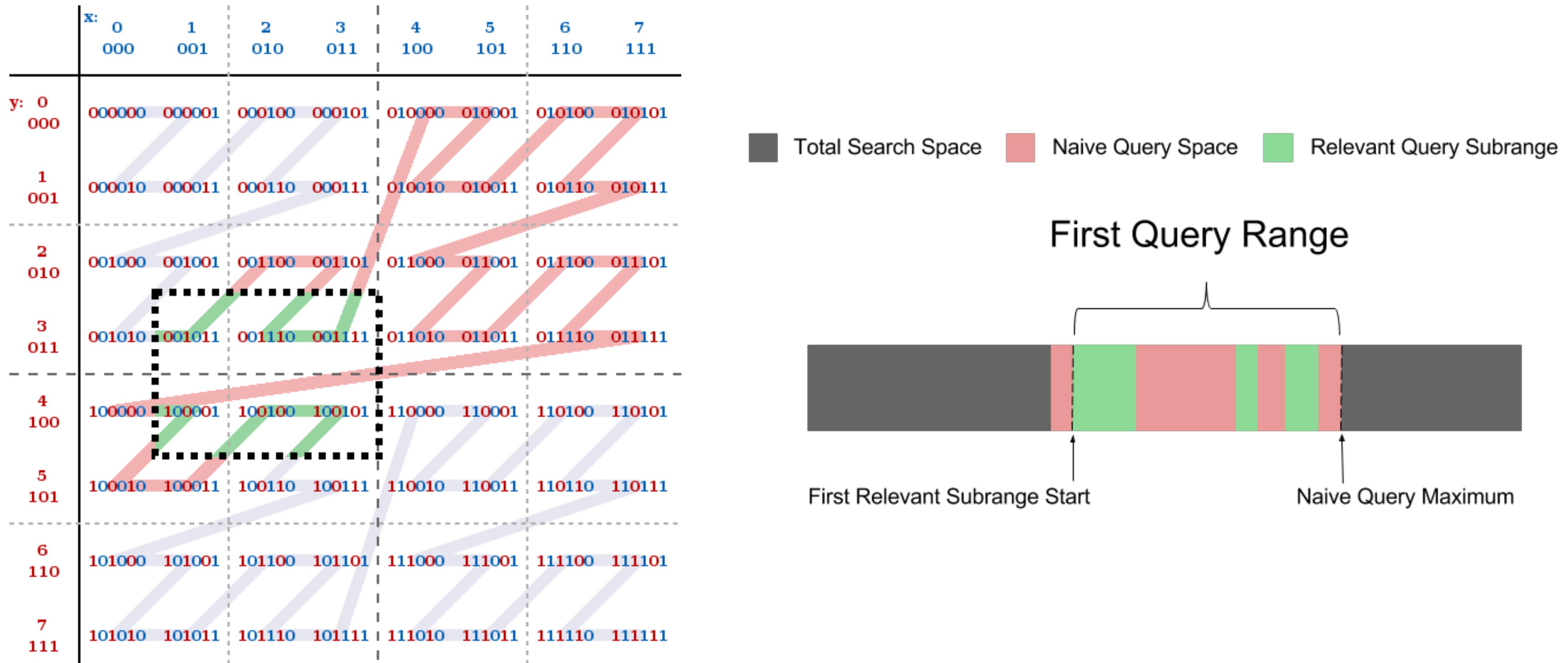
1 1 0 0 1

1101001001

1 0 0 1 0

Существует 10 типов людей: те, кто понимает двоичную систему счисления, и те, кто не понимают.

Z-order curve



Пример

```
CREATE TABLE test_zcurve_two AS datasets.hits_v1
ENGINE = MergeTree
ORDER BY zCurve(CounterID, UserID)
```

ok.

```
CREATE TABLE test_tuple_two AS datasets.hits_v1
ENGINE = MergeTree
ORDER BY (CounterID, UserID)
```

ok.

Пример

```
khvostikovao-nix :) select count() from test_zcurve_two where CounterID > 158585 and UserID < 2007878526839192;  
  
SELECT count()  
FROM test_zcurve_two  
WHERE (CounterID > 158585) AND (UserID < 2007878526839192)
```

```
count()  
15438
```

1 rows in set. Elapsed: 0.045 sec. Processed 499.71 thousand rows, 5.73 MB (11.11 million rows/s., 127.42 MB/s.)

```
khvostikovao-nix :) select count() from test_tuple_two where CounterID > 158585 and UserID < 2007878526839192;
```

```
SELECT count()  
FROM test_tuple_two  
WHERE (CounterID > 158585) AND (UserID < 2007878526839192)
```

```
count()  
15438
```

1 rows in set. Elapsed: 0.074 sec. Processed 5.24 million rows, 42.02 MB (71.17 million rows/s., 570.19 MB/s.)

Дополнительные фичи

Многомерность

Нормализация



Спасибо

Ольга Хвостикова
Разработчик



На сегодня всё.
Расходимся.

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

$$\sum_{n=1}^{\infty} \frac{n}{7n+3}$$

$$\sum_{n=1}^{\infty} \frac{n-1}{\sqrt{2n^4-n+5}}$$

$$\sum_{n=2}^{\infty} \frac{1}{\ln n}$$

$$\sum_{n=1}^{\infty} \frac{n^3 - 3n + 1}{n^2 + 4}$$

$$\sum_{n=1}^{\infty} \frac{1}{\sin n}$$

pikabu.ru