



**BIGDATA
TEAM**

Тестирование приложений, pytest

Реальная жизнь vs оптимизация, δ-coding, TF-IDF, PageRank и КОСТЫЛИ

Dral Alexey, study@bigdatateam.org

CEO at BigData Team, <https://bigdatateam.org>

<https://www.facebook.com/bigdatateam>



**BIGDATA
TEAM**

Реальная жизнь



Сложные алгоритмы



Бизнес-ориентированные решения



Костыли



Золотое правило программирования



совершенство достигнуто не тогда,
когда нечего добавить, а тогда, когда
нечего убрать ©

Антуан Мари Жан-Батист Рожэ де Сент-Экзюпері
29 июня 1900, Лион, Франция — 31 июля 1944



Кодирование Timestamp

2017-05-19 17:53

2017-05-19 17:59

2017-05-19 18:02

2017-05-20 10:27

...

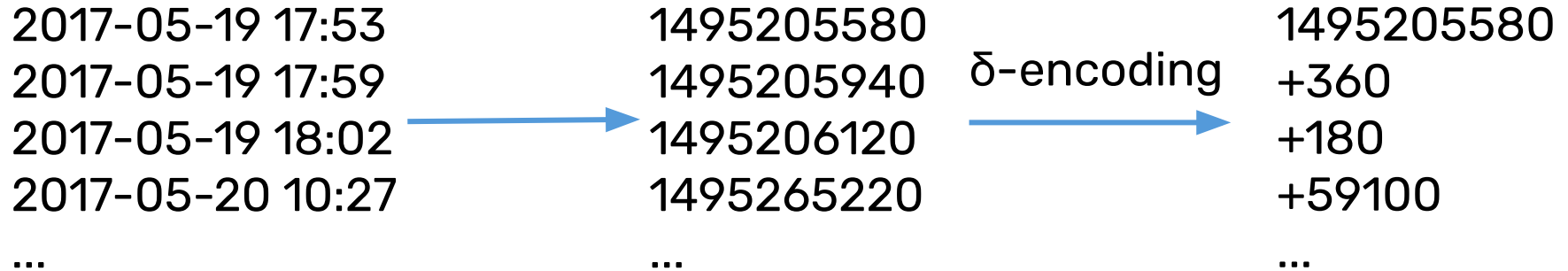


Кодирование Timestamp

2017-05-19 17:53	1495205580
2017-05-19 17:59	1495205940
2017-05-19 18:02	1495206120
2017-05-20 10:27	1495265220
...	...



Кодирование Timestamp





$$W_{x,y} = tf_{x,y} \times \log(N / df_x)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

