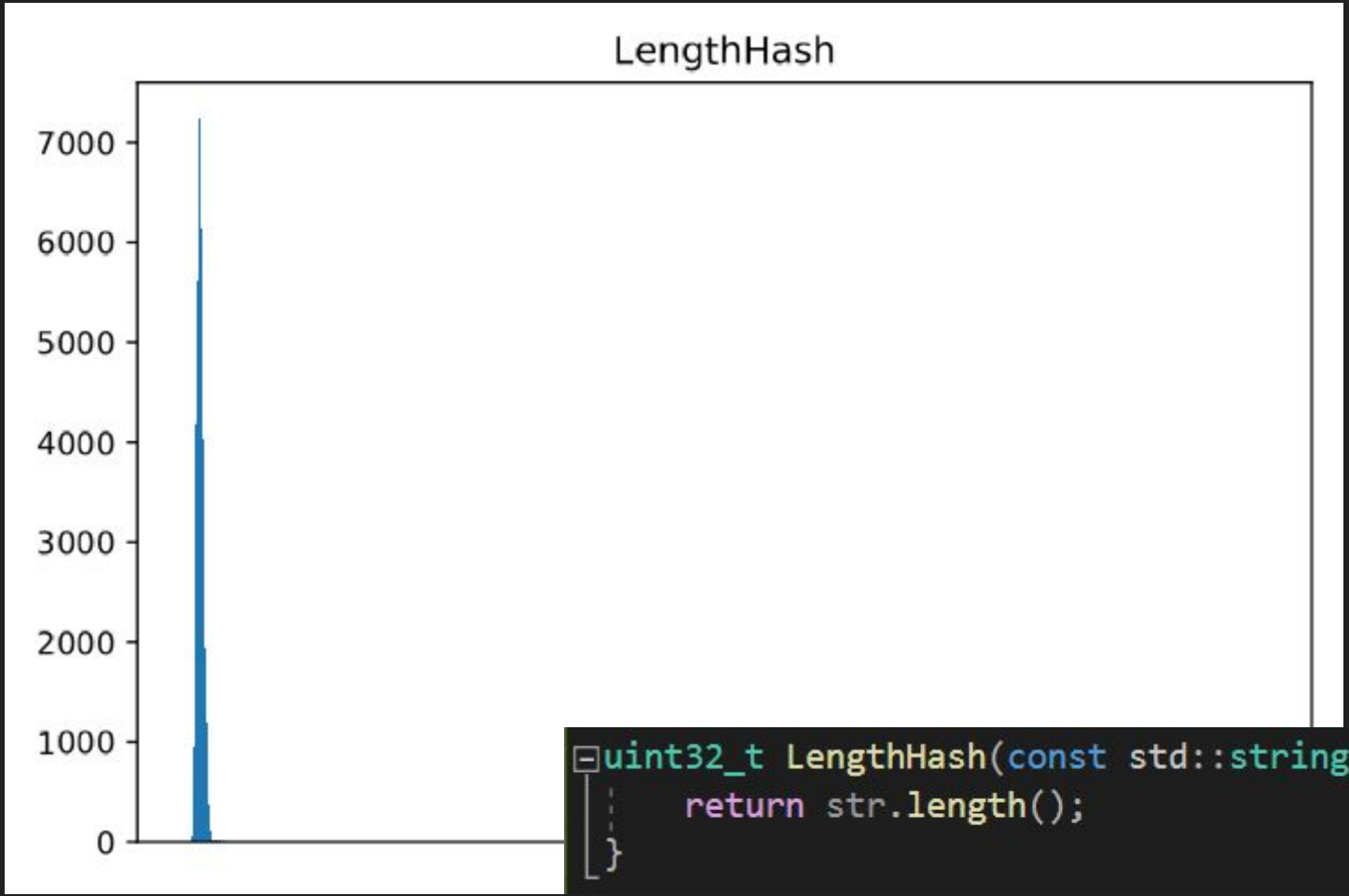
Оптимизация хеш-таблицы

Демина Елизавета, 2020

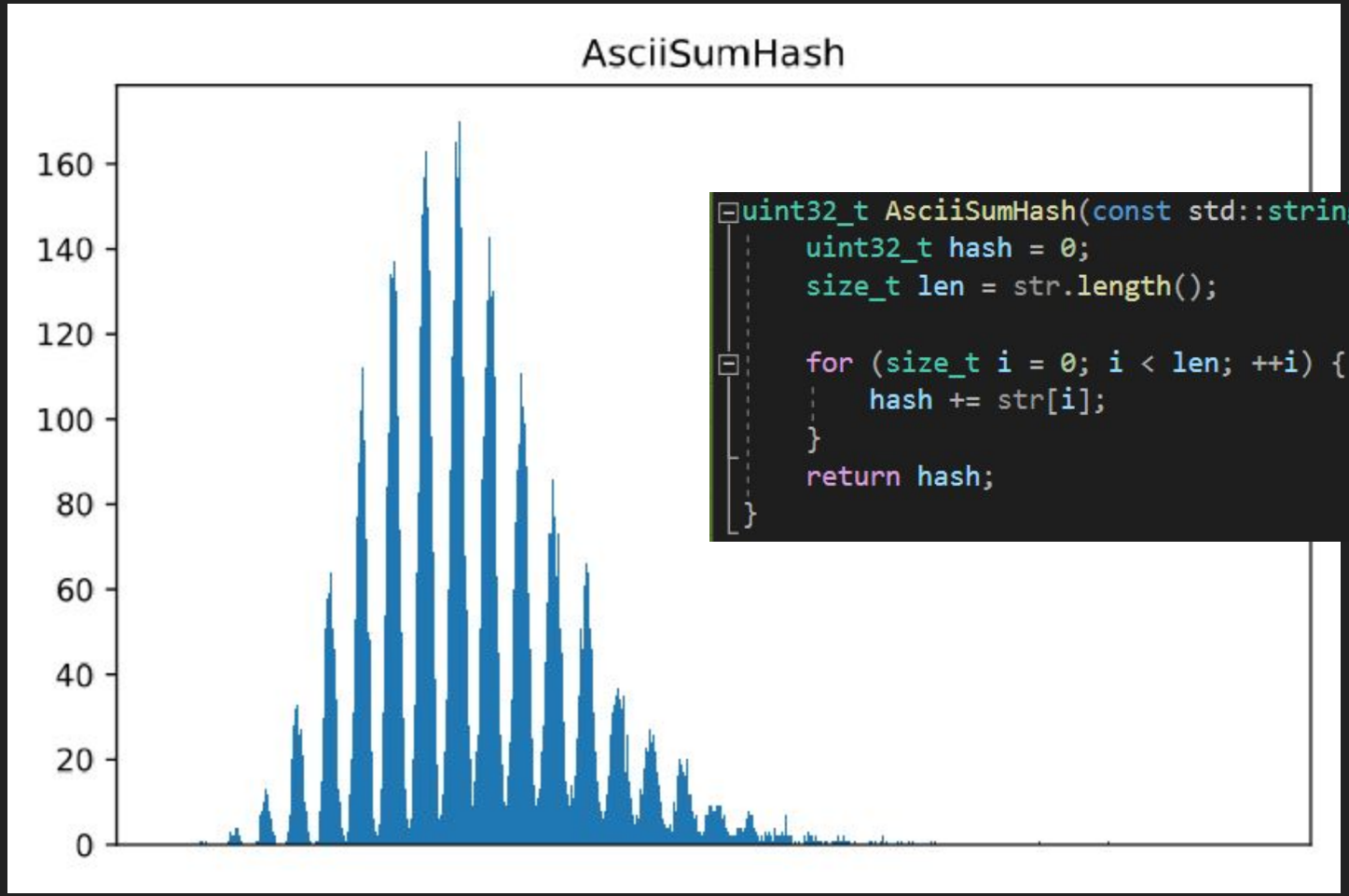
Хеш-таблица

- 50.000 английских слов
- Слова длиной от 2 до 30 символов
- 6 различных хеш-функций
- Размер таблицы: 3671


```
uint32_t Hash1(const std::string& str) {
    return 1;
}
```

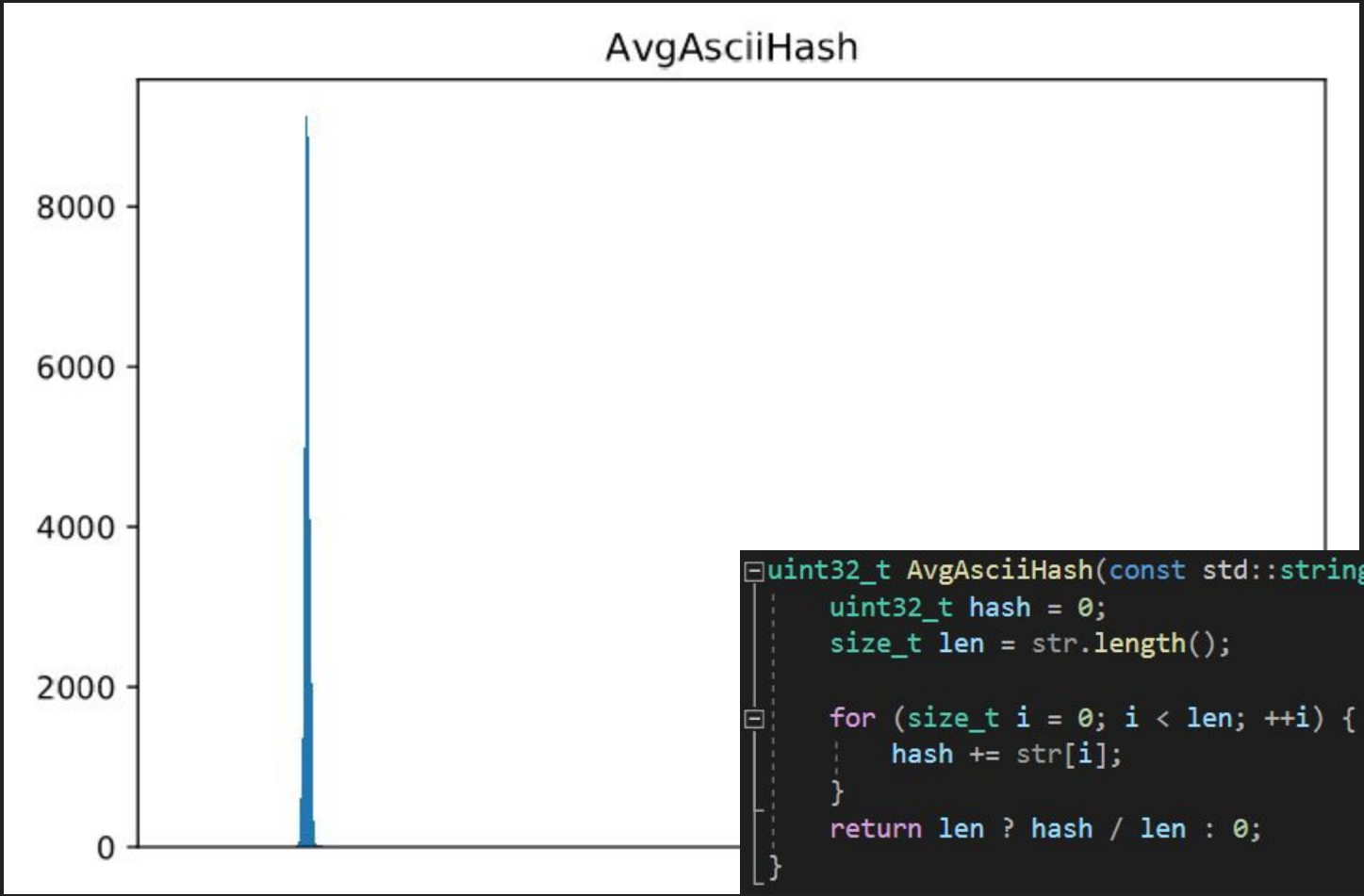



```cpp
uint32_t LengthHash(const std::string& str) {
    return str.length();
}
```



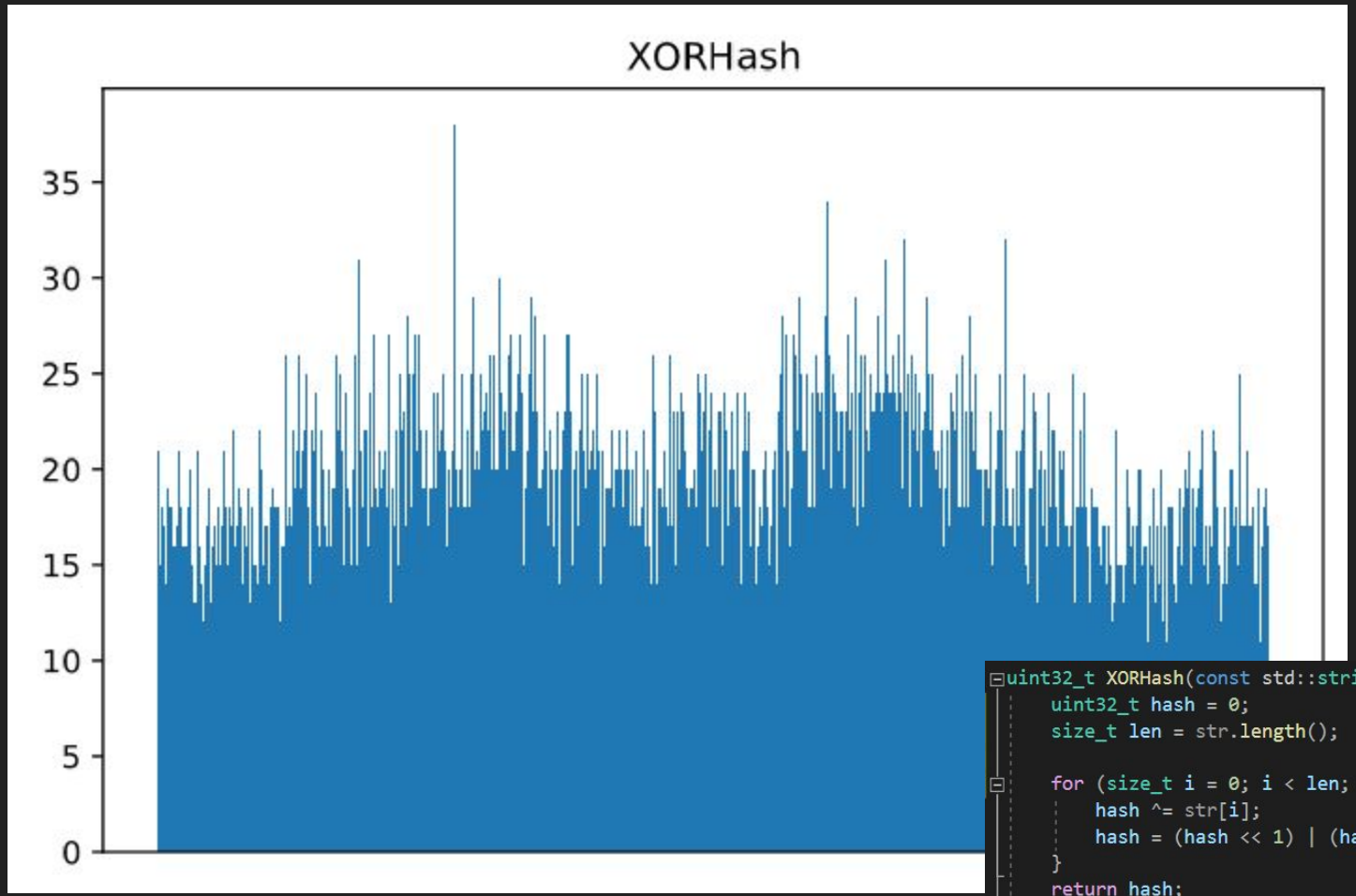

```cpp
uint32_t AsciiSumHash(const std::string& str) {
    uint32_t hash = 0;
    size_t len = str.length();

    for (size_t i = 0; i < len; ++i) {
        hash += str[i];
    }
    return hash;
}
```



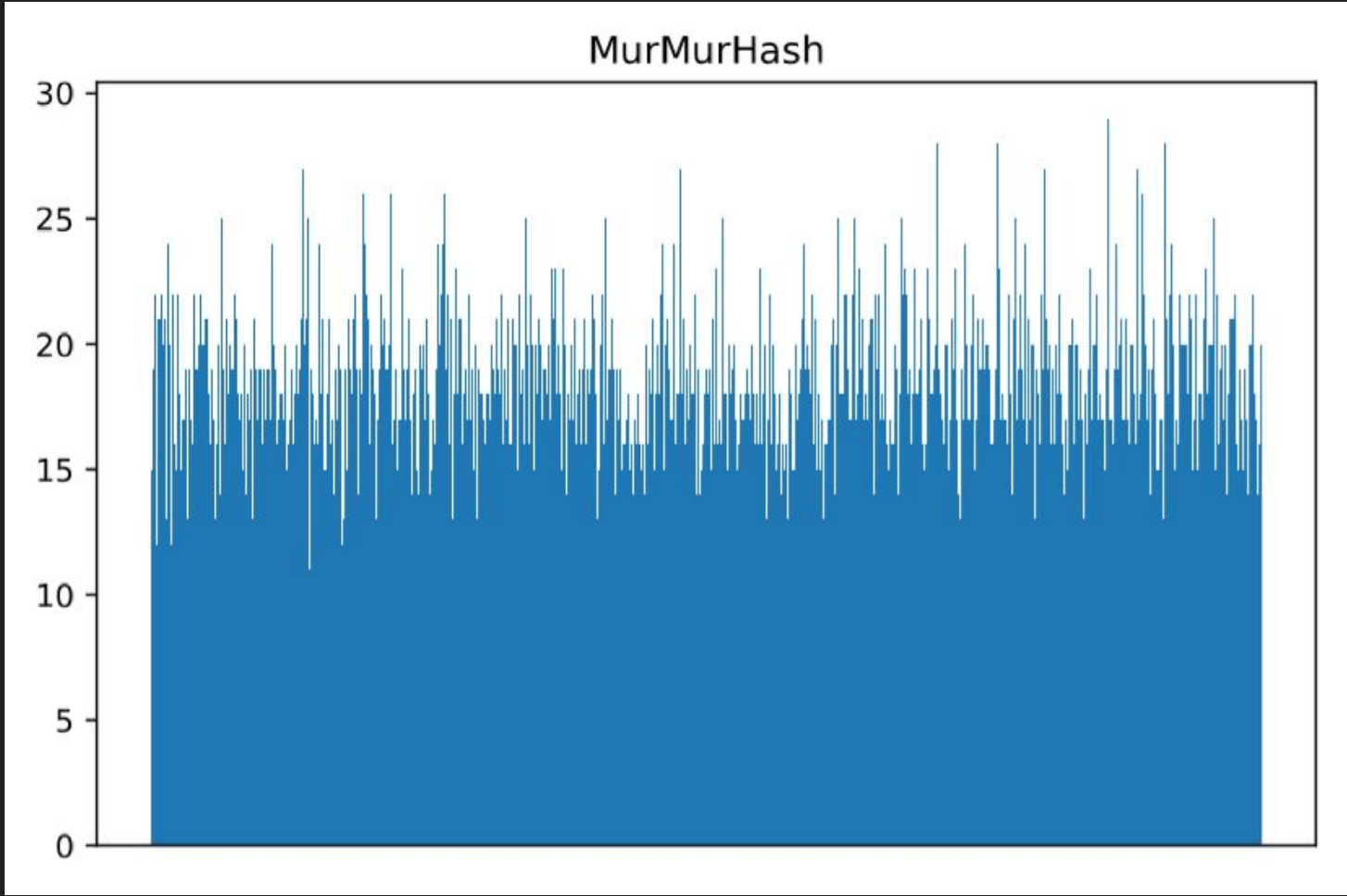

```cpp
uint32_t AvgAsciiHash(const std::string& str) {
    uint32_t hash = 0;
    size_t len = str.length();

    for (size_t i = 0; i < len; ++i) {
        hash += str[i];
    }
    return len ? hash / len : 0;
}
```




```cpp
uint32_t XORHash(const std::string& str) {
    uint32_t hash = 0;
    size_t len = str.length();

    for (size_t i = 0; i < len; ++i) {
        hash ^= str[i];
        hash = (hash << 1) | (hash >> 31);
    }
    return hash;
}
```


MurMurHash
30
25
20
15
10
5
0

Профилирование хеш-таблицы

std::_Default_allocator_traits<std::allocator<std::_...	3160 (5,66%)	154 (0,28%)	hash_table1.exe
std::_Default_allocator_traits<std::allocator<std::_...	3040 (5,45%)	154 (0,28%)	hash_table1.exe
std::_List_node_emplace_op2<std::allocator<std::_...	2759 (4,94%)	708 (1,27%)	hash_table1.exe
std::basic_string<char,std::char_traits<char>,std::a...	2576 (4,61%)	248 (0,44%)	hash_table1.exe
std::basic_string<char,std::char_traits<char>,std::a...	2363 (4,23%)	485 (0,87%)	hash_table1.exe
XORHash	2292 (4,11%)	175 (0,31%)	hash_table1.exe
AvgAsciiHash	2269 (4,06%)	131 (0,23%)	hash_table1.exe
AsciiSumHash	2160 (3,87%)	133 (0,24%)	hash_table1.exe
std::_String_val<std::_Simple_types<char> >::_Lar...	2056 (3,68%)	1865 (3,34%)	hash_table1.exe
std::_Construct_in_place<std::_List_node<std::basi...	1913 (3,43%)	498 (0,89%)	hash_table1.exe
std::basic_string<char,std::char_traits<char>,std::a...	1734 (3,11%)	358 (0,64%)	hash_table1.exe
std::_Unfancy<char>	1688 (3,02%)	1326 (2,38%)	hash_table1.exe
std::_Container_base12::_Orphan_all	1579 (2,83%)	1024 (1,83%)	hash_table1.exe

Реализации XORHash

```cpp
uint32_t XORHash(const std::string& str) {
    uint32_t hash = 0;
    size_t len = str.length();

    for (size_t i = 0; i < len; ++i) {
        hash ^= str[i];
        hash = (hash << 1) | (hash >> 31);
    }
    return hash;
}
```

C++

```asm
_XORHashAsm proc string:ptr byte
    mov edx, dword ptr string
    xor eax, eax            ; eax = 0
    xor ebx, ebx            ; ebx = 0
    hash_loop:
        mov bl, [edx]       ; bl = str[i]
        cmp bl, 0h
        je exit
        xor eax, ebx        ; eax ^= ebx
        rol eax, 1          ; eax circular shift right 1
        inc edx             ; edx++
        jmp hash_loop
    exit:
        ret
_XORHashAsm endp
```

Asm

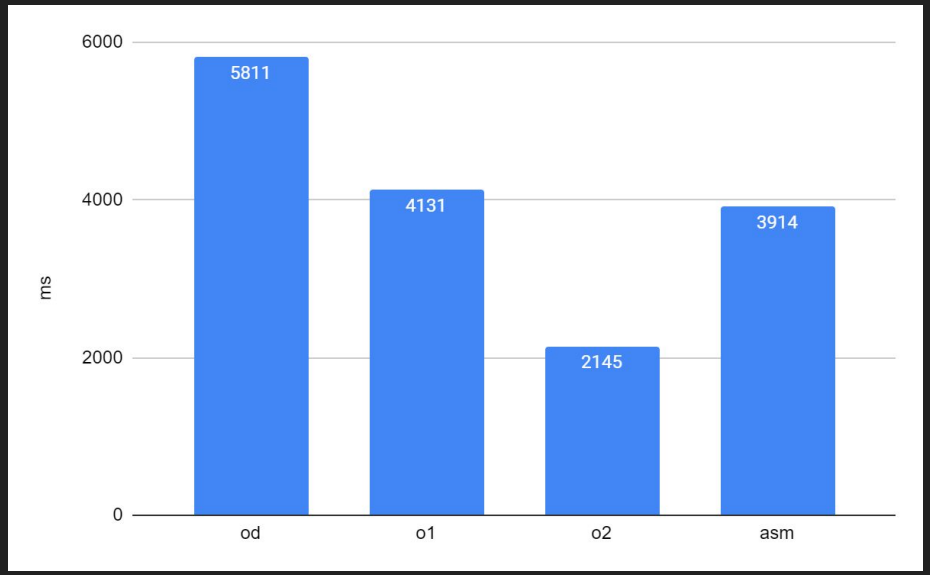
Было

main	18774 (98,21%)
TimeTest<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >	17956 (93,93%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	9749 (51,00%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Contains	7738 (40,48%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	7452 (38,98%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Contains	5318 (27,82%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Find	5219 (27,30%)
std::operator==<char,std::char_traits<char>,std::allocator<char> >	4782 (25,02%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::GetIndex	4390 (22,97%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::_Equal	4335 (22,68%)
operator new	4322 (22,61%)
[Внешний вызов] malloc	4293 (22,46%)
XORHash	4008 (20,97%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::Node	3457 (18,08%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::basic_string<char,std::c...	3335 (17,45%)
__CheckForDebuggerJustMyCode	2999 (15,69%)
std::_Container_base12::_Alloc_proxy<std::allocator<std::_Container_proxy> >	2756 (14,42%)
std::_String_val<std::_Simple_types<char> >::_Myptr	2744 (14,35%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::operator[]	2593 (13,56%)
std::allocator<std::_Container_proxy>::allocate	2301 (12,04%)
std::_Allocate<8,std::_Default_allocate_traits,0>	2267 (11,86%)

Стало

std::basic_string<char,std::char_traits<char>,std::allocator<char> >::_Equal	5143 (26,38%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::Node	4432 (22,73%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::basic_string<char,std::c...	4302 (22,06%)
std::_Container_base12::_Alloc_proxy<std::allocator<std::_Container_proxy> >	3529 (18,10%)
std::allocator<std::_Container_proxy>::allocate	2903 (14,89%)
std::_Allocate<8,std::_Default_allocate_traits,0>	2868 (14,71%)
std::_Default_allocate_traits::_Allocate	2790 (14,31%)
__CheckForDebuggerJustMyCode	2693 (13,81%)
std::_Traits_equal<std::char_traits<char> >	2047 (10,50%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::operator=	1870 (9,59%)
std::_String_val<std::_Simple_types<char> >::_Myptr	1816 (9,31%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::_Copy_assign	1748 (8,97%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::GetIndex	1271 (6,52%)
std::_Narrow_char_traits<char,int>::compare	1239 (6,35%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::assign	1191 (6,11%)
[Внешний вызов] memcmp	1075 (5,51%)
std::_String_val<std::_Simple_types<char> >::_Large_string_engaged	944 (4,84%)
XORHashAsm	826 (4,24%)
ReadData	505 (2,59%)
std::_Construct_in_place<std::_Container_proxy,std::_Container_base12 *>	406 (2,08%)

Замеры времени



	Od	O1	O2	Asm
Ср. время работы (мс)	5811	4131	2145	3914
Коэфф.ускорения	1.5	1.1	0.5	1