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

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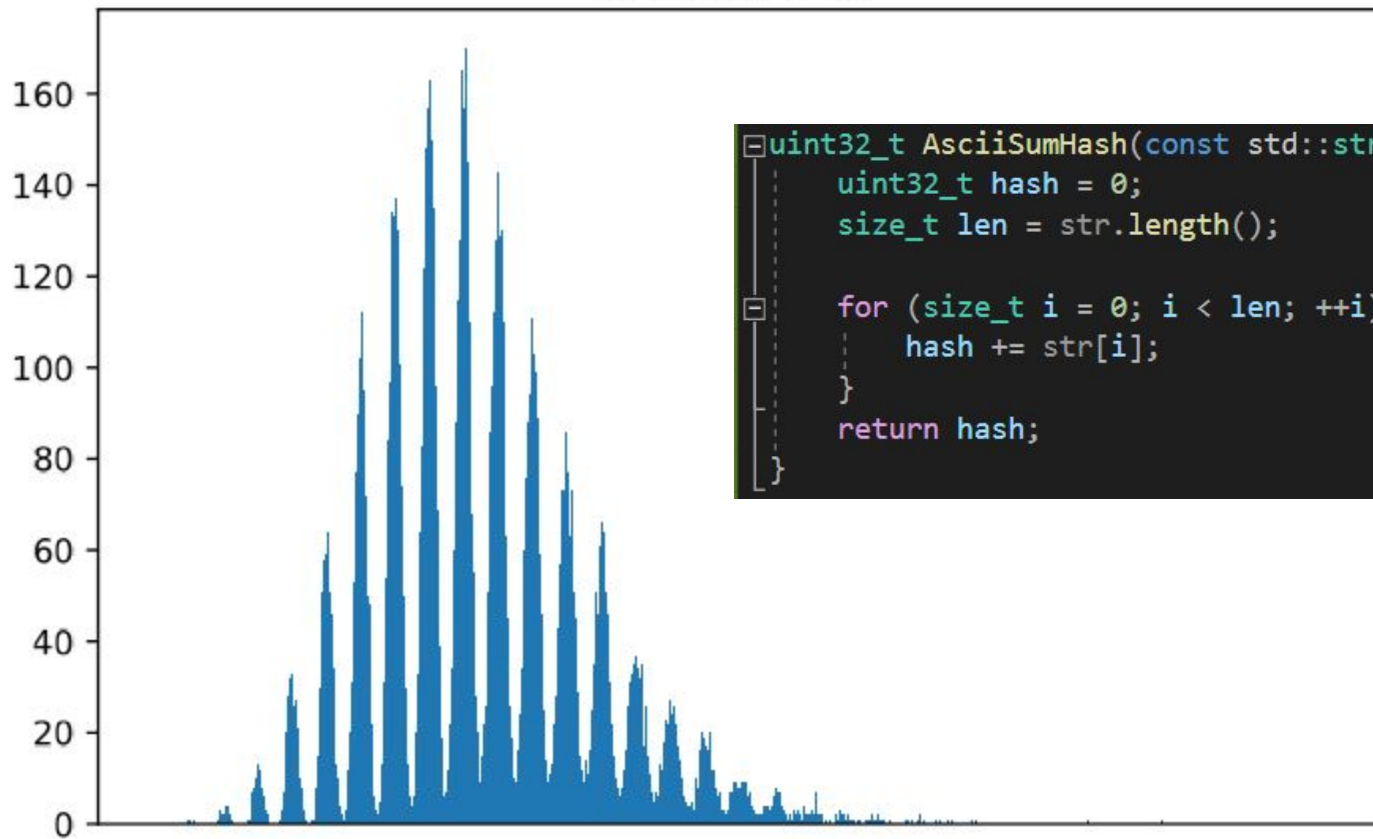
```cpp
uint32_t Hash1(const std::string& str) {
    return 1;
}
```
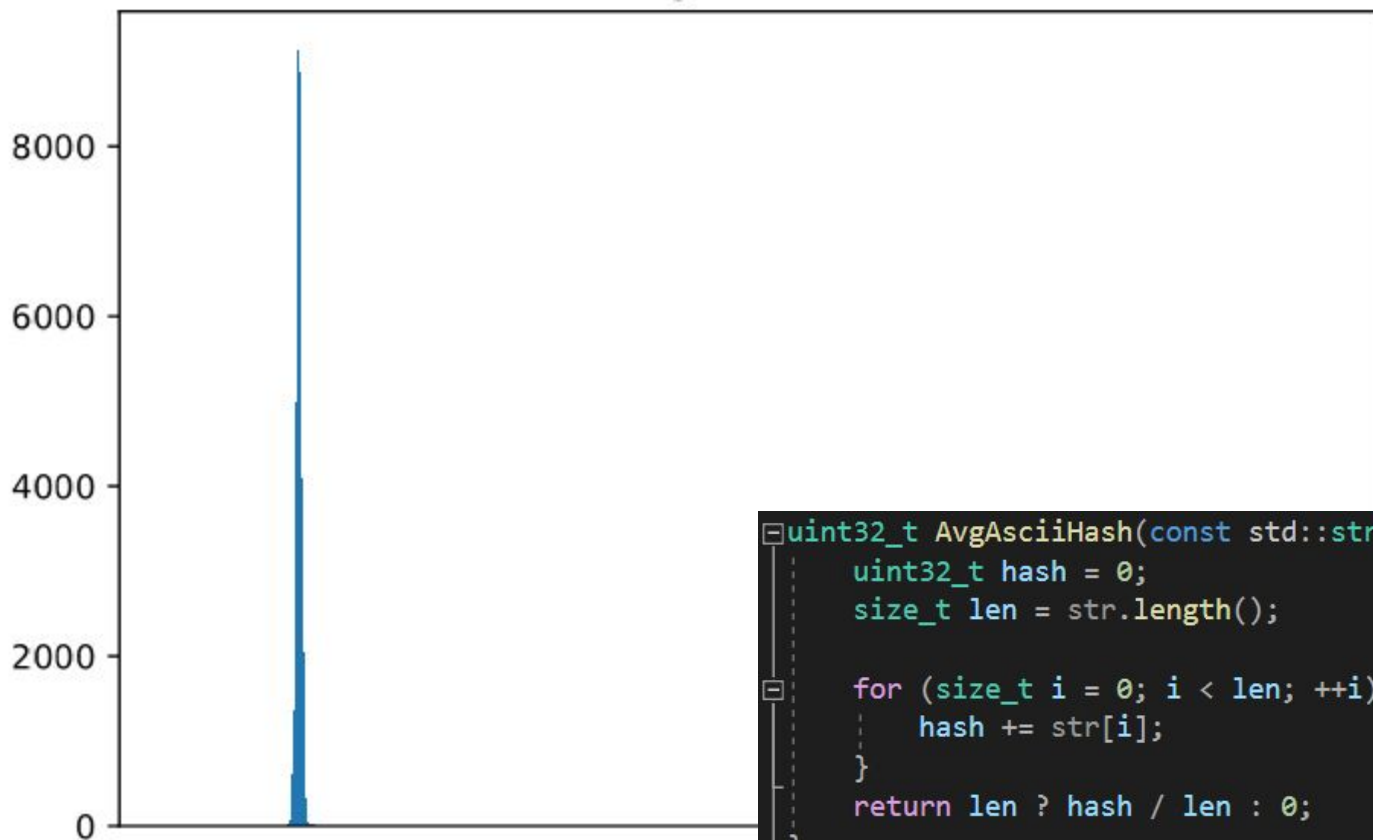
LengthHash


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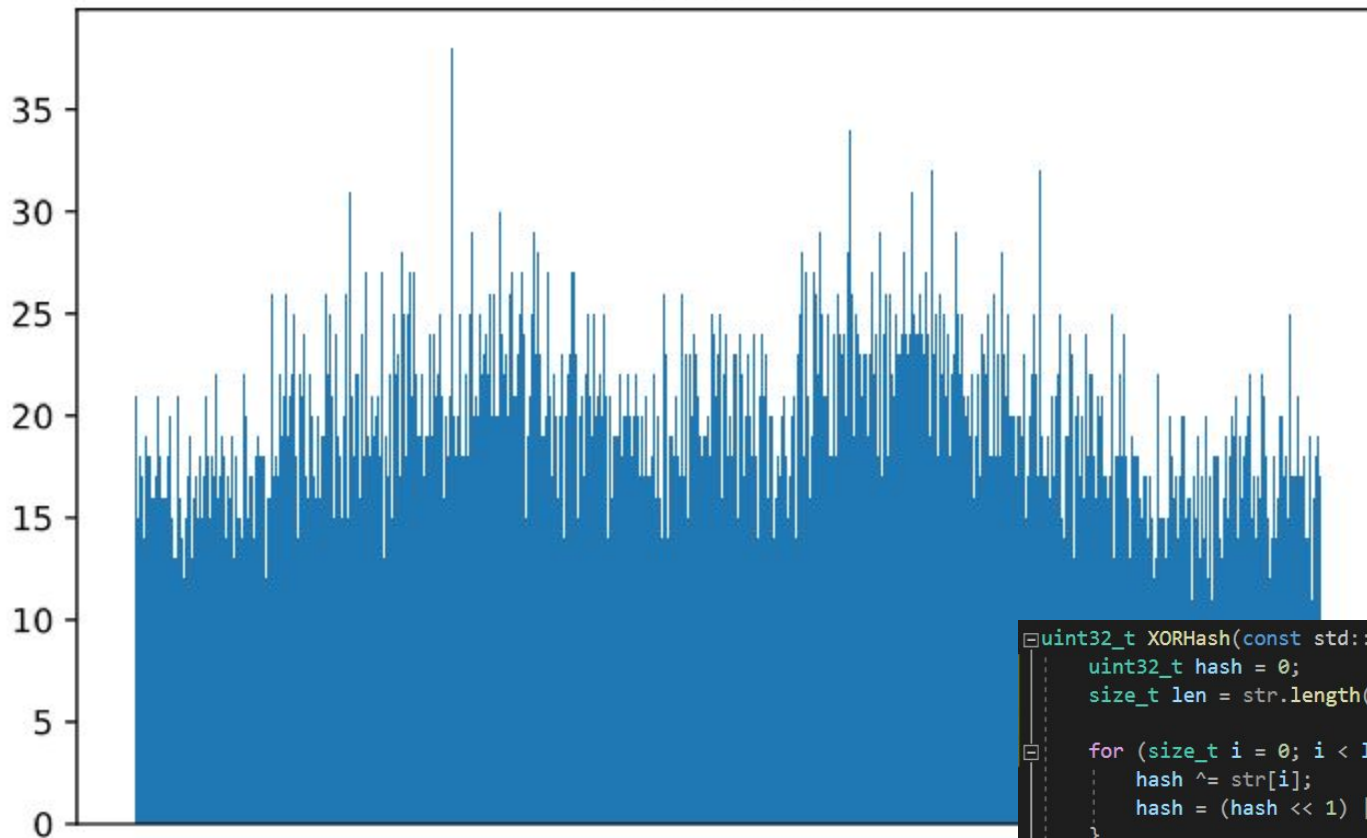
AsciiSumHash

```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;
}
```

AvgAsciiHash


```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;
}
```
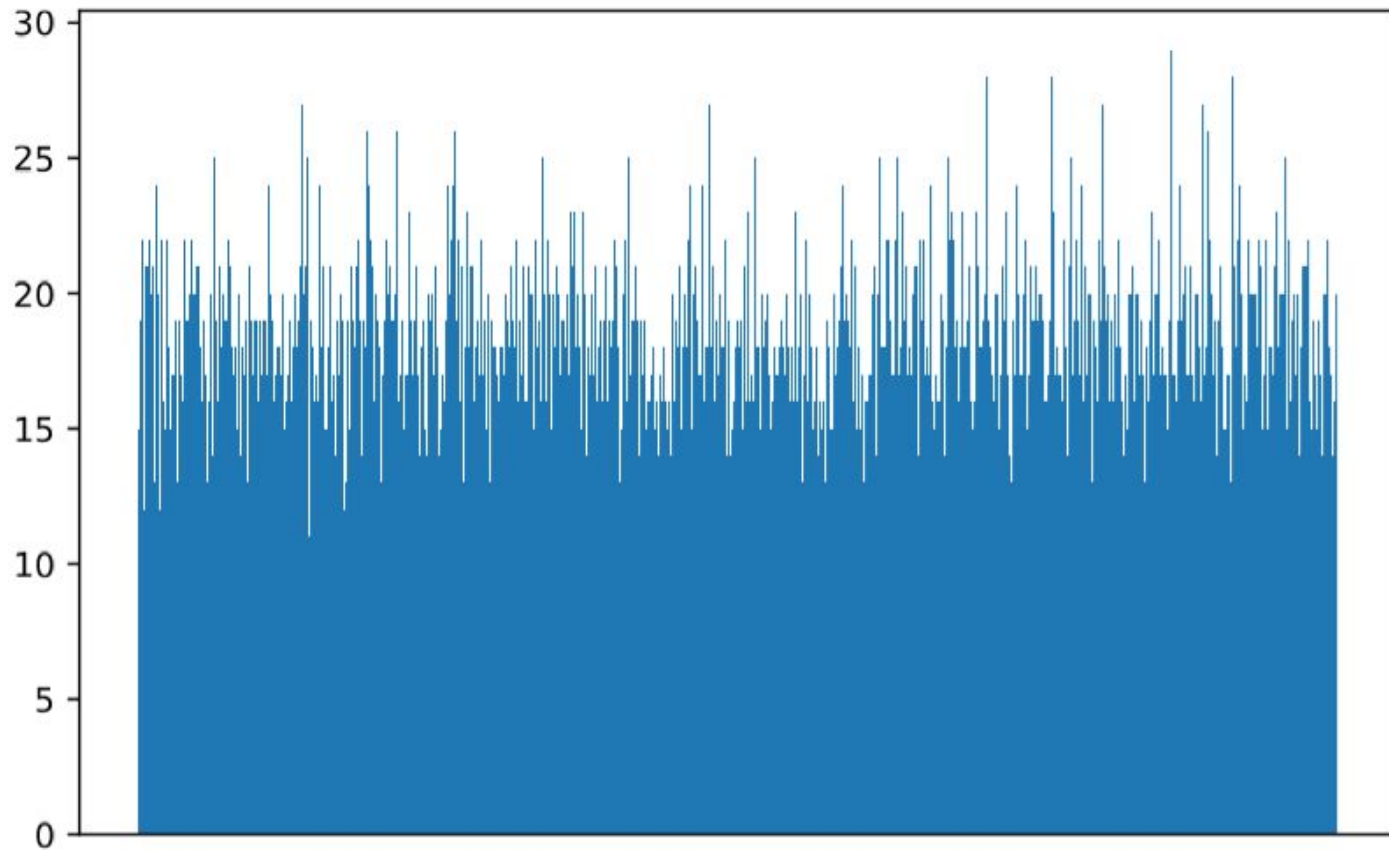
XORHash


```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;
}
```

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;
}
```

```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
```
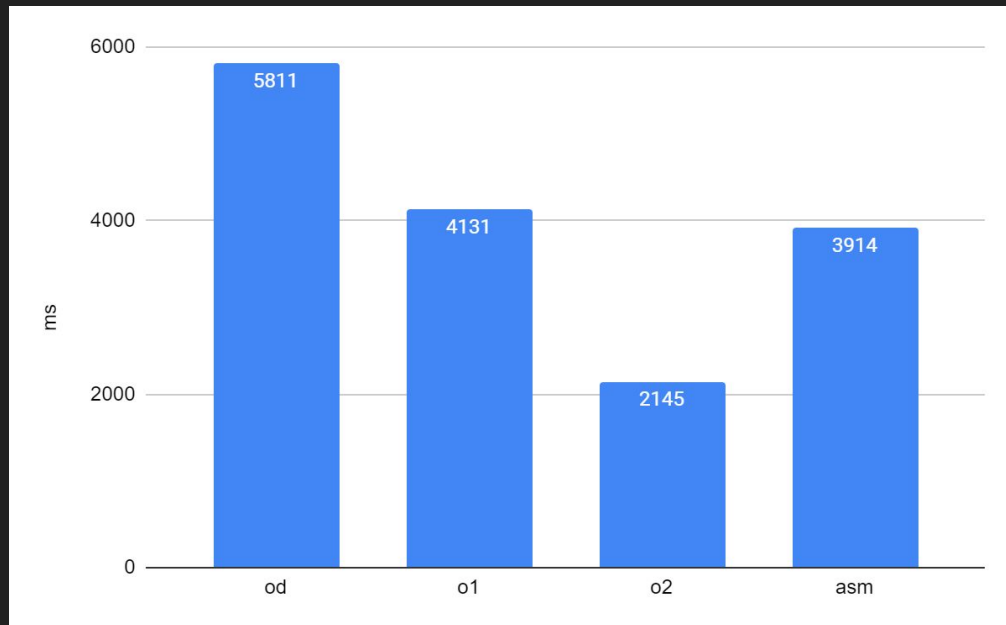
C++   Asm

Было

[Внешний код]	10926 (99,99%)	5406 (49,47%)
__scrt_common_main	10852 (99,31%)	0 (0,00%)
__scrt_common_main_seh	10852 (99,31%)	0 (0,00%)
mainCRTStartup	10852 (99,31%)	0 (0,00%)
invoke_main	10849 (99,29%)	0 (0,00%)
main	10849 (99,29%)	0 (0,00%)
TimeTest<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >	5646 (51,67%)	15 (0,14%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::~HashTa...	4893 (44,78%)	0 (0,00%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::`vector deleting...	4893 (44,78%)	0 (0,00%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::~List<std::basic_...	4892 (44,77%)	1 (0,01%)
`eh vector destructor iterator'	4892 (44,77%)	0 (0,00%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::RemoveHead	4886 (44,71%)	636 (5,82%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::`scalar del...	4239 (38,79%)	12 (0,11%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	4035 (36,93%)	15 (0,14%)
operator delete	3086 (28,24%)	10 (0,09%)
operator delete	3074 (28,13%)	16 (0,15%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	2947 (26,97%)	147 (1,35%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::~basic_string<char,std::...	2818 (25,79%)	25 (0,23%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::~Node	2781 (25,45%)	8 (0,07%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::GetIndex	2078 (19,02%)	123 (1,13%)
operator new	1980 (18,12%)	21 (0,19%)
XORHash	1899 (17,38%)	568 (5,20%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::Node	1758 (16,09%)	19 (0,17%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::basic_string<char,std::c...	1735 (15,88%)	52 (0,48%)
std::_Delete_plain_internal<std::allocator<std::_Container_proxy> >	1690 (15,47%)	18 (0,16%)

Стало

List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::`scalar del...	4253 (44,52%)	22 (0,23%)
TimeTest<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >	4222 (44,19%)	8 (0,08%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	3314 (34,69%)	12 (0,13%)
operator delete	3076 (32,20%)	14 (0,15%)
operator delete	3050 (31,92%)	11 (0,12%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Insert	2993 (31,33%)	105 (1,10%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::~basic_string<char,std:...	2919 (30,55%)	42 (0,44%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::~Node	2880 (30,14%)	14 (0,15%)
operator new	2041 (21,36%)	18 (0,19%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Node::Node	1858 (19,45%)	17 (0,18%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::basic_string<char,std::c...	1832 (19,18%)	47 (0,49%)
std::_Deallocate<8,0>	1747 (18,29%)	14 (0,15%)
std::_Delete_plain_internal<std::allocator<std::_Container_proxy> >	1730 (18,11%)	10 (0,10%)
std::_Deallocate_plain<std::allocator<std::_Container_proxy> >	1719 (17,99%)	12 (0,13%)
std::_Default_allocator_traits<std::allocator<std::_Container_proxy> >::deallocate	1696 (17,75%)	5 (0,05%)
std::_Container_proxy_ptr12<std::allocator<std::_Container_proxy> >::_Container_proxy_ptr...	1241 (12,99%)	42 (0,44%)
std::_Allocate<8,std::_Default_allocate_traits,0>	1050 (10,99%)	9 (0,09%)
std::allocator<std::_Container_proxy>::allocate	1049 (10,98%)	11 (0,12%)
std::_Default_allocate_traits::_Allocate	1035 (10,83%)	8 (0,08%)
std::basic_string<char,std::char_traits<char>,std::allocator<char> >::_Tidy_deallocate	1027 (10,75%)	26 (0,27%)
std::_Container_base12::_Orphan_all	897 (9,39%)	737 (7,71%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Contains	737 (7,71%)	20 (0,21%)
__CheckForDebuggerJustMyCode	602 (6,30%)	600 (6,28%)
HashTable<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::GetIndex	587 (6,14%)	128 (1,34%)
XORHashAsm	409 (4,28%)	39 (0,41%)
List<std::basic_string<char,std::char_traits<char>,std::allocator<char> > >::Contains	406 (4,25%)	27 (0,28%)

Замеры времени

	Od	O1	O2	Asm
Тест 1 (ms)	5842	4117	2140	3915
Тест 2 (ms)	5768	4071	2159	3926
Тест 3 (ms)	5793	4277	2142	3899
Тест 4 (ms)	5788	4134	2151	3904
Тест 5 (ms)	5864	4052	2134	3924
Среднее	5811	4131	2145	3914
Коэфф.уск.	1.5	1.1	0.5	1