

Национальный исследовательский университет
«Высшая Школа Экономики»
Факультет компьютерных наук

«С++ как первый язык программирования»

Лекция 14

Лекторы: Роман Халкечев, Кирилл Лунев

Цели лекции и семинара

- Поговорим про итераторы.
- Напишем программы построения обратного индекса по коллекции документов и поиска документов по запросу пользователя.

Итераторы

- Итератор (от англ. iterator — перечислитель) — интерфейс, предоставляющий доступ к элементам коллекции (массива или контейнера) и навигацию по ним.
- Итераторы имеются у всех контейнеров STL. Методы (begin(), end(), rbegin(), rend(), ...).
- Большинство алгоритмов STL для работы с контейнерами принимают на вход и возвращают итераторы.

Примеры алгоритмов STL

- [std::find](#) — алгоритм поиска элемента в контейнере.
- [std::count](#) — подсчёт числа элементов равных заданном в контейнере.
- [std::copy](#) — копирование диапазона элементов контейнера.
- [std::copy_if](#) — копирование элементов из диапазона контейнера при условии выполнения некоторого условия.
- [std::sort](#) — сортировка диапазона элементов контейнера.
- ...

Inverted index

Инвертированный индекс (англ. inverted index) — структура данных, в которой для каждого слова коллекции документов в соответствующем списке перечислены все документы в коллекции, в которых оно встретилось.

Document 1

The bright blue butterfly hangs on the breeze.

Document 2

It's best to forget the great sky and to retire from every wind.

Document 3

Under blue sky, in bright sunlight, one need not search around.

Stopword list

a
and
around
every
for
from
in
is
it
not
on
one
the
to
under

Inverted index

ID	Term	Document
1	best	2
2	blue	1, 3
3	bright	1, 3
4	butterfly	1
5	breeze	1
6	forget	2
7	great	2
8	hangs	1
9	need	3
10	retire	2
11	search	3
12	sky	2, 3
13	wind	2

Document 1

The bright blue butterfly hangs on the breeze.

Document 2

It's best to forget the great sky and to retire from every wind.

Document 3

Under blue sky, in bright sunlight, one need not search around.

Query

Q- "blue sky"

Inverted index

ID	Term	Document : position
1	best	2 : 3
2	blue	1 : 3, 3 : 2
3	bright	1 : 2, 3 : 5
4	butterfly	1 : 4
5	breeze	1 : 8
6	forget	2 : 5
7	great	2 : 7
8	hangs	1 : 5
9	needs	3 : 8
10	retire	2 : 11
11	search	3 : 10
12	sky	2 : 8, 3 : 3
13	wind	2 : 14

Match on sequential terms

blue - 3 : 2
sky - 3 : 3

Search object

Document reference	Relevance
3	100%

Задачи

- **Задача 1:** Напишите программу, которая строит обратный индекс по предоставленной коллекции текстовых документов. Коллекцию документов можно взять по ссылке: <https://yadi.sk/d/P6G79Q7VsbTFx>
- **Задача 2:** Напишите программу, которая осуществляет булев поиск по построенному индексу из первой задачи. Программа считывает запросы со стандартного потока ввода по одному на строке, а затем выводит первые 10 документов, содержащих все слова запроса, плюс общее число найденных документов.

Вопросы?

Домашнее задание

- Дорешать задачи