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

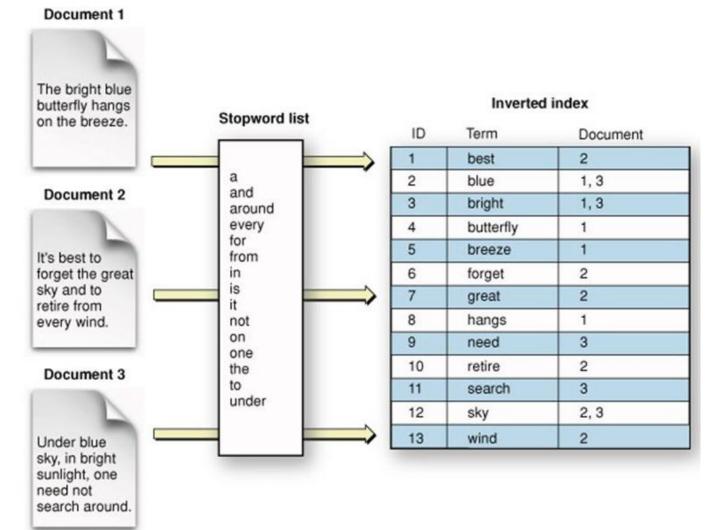
- Итератор (от англ. iterator перечислитель) интерфейс,
 предоставляющий доступ к элементам коллекции (массива или контейнера) и навигацию по ним.
- Итераторы имеются у всех контейнеров STL. Методы (begin(), end(), rbegin(), rend(), ...).
- Большинство алгоритмов STL для работы с контейнерами принимают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

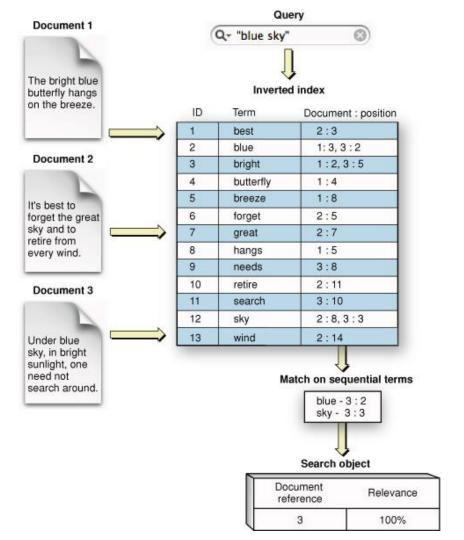
Примеры алгоритмов STL

- <u>std::find</u> алгоритм поиска элемента в контейнере.
- <u>std::count</u> подсчёт числа элементов равных заданном в контейнере.
- <u>std::copy</u> копирование диапазона элементов контейнера.
- std::copy_if копирование элементов из диапазона контейнера при условии выполнения некоторого условия.
- <u>std::sort</u> 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.

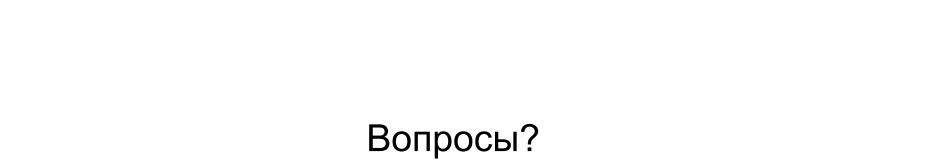




Задачи

• Задача 1: Напишите программу, которая строит обратный индекс по предоставленной коллекции текстовых документов. Колекцию 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.



Домашнее задание

• Дорешать задачи