

CSCI-1200 Data Structures — Fall 2016

Lecture 23 – Functors & Hash Tables, part II

Review from Lecture 22

- Hash Tables, Hash Functions, and Collision Resolution
- Performance of: Hash Tables vs. Binary Search Trees
- Collision resolution: separate chaining vs open addressing
- STL's `unordered_set` (and `unordered_map`)

Today's Lecture

- Using STL's `for_each`
- Something weird & cool in C++... Function Objects, a.k.a. *Functors*
- Continuing with Hash Tables...
 - STL's `unordered_set` (and `unordered_map`)
 - Using a hash table to implement a set/map
 - Hash functions as functors/function objects
 - Iterators, find, insert, and erase
- Homework 9 Preview & Discussion

23.1 Using STL's `for_each`

- First, here's a tiny helper function:

```
void float_print (float f) {  
    std::cout << f << std::endl;  
}
```

- Let's make an STL vector of floats:

```
std::vector<float> my_data;  
my_data.push_back(3.14);  
my_data.push_back(1.41);  
my_data.push_back(6.02);  
my_data.push_back(2.71);
```

- Now we can write a loop to print out all the data in our vector:

```
std::vector<float>::iterator itr;  
for (itr = my_data.begin(); itr != my_data.end(); itr++) {  
    float_print(*itr);  
}
```

- Alternatively we can use it with STL's `for_each` function to visit and print each element:

```
std::for_each(my_data.begin(), my_data.end(), float_print);
```

Wow! That's alot less to type. Can I stop using regular `for` and `while` loops altogether?

- We can actually also do the same thing without creating & explicitly naming the `float_print` function. We create an *anonymous function* using *lambda*:

```
std::for_each(my_data.begin(), my_data.end(), [](float f){ std::cout << f << std::endl; });
```

Lambda is new to the C++ language (part of C++11). But lambda is a core piece of many classic, older programming languages including Lisp and Scheme. Python lambdas and Perl anonymous subroutines are similar. (In fact lambda dates back to the 1930's, before the first computers were built!) You'll learn more about lambda more in later courses like CSCI 4430 Programming Languages!

23.2 Function Objects, a.k.a. *Functors*

- In addition to the basic mathematical operators + - * / < >, another operator we can overload for our C++ classes is the *function call operator*.
Why do we want to do this? This allows instances or objects of our class, to be used like functions. It's weird but powerful.
- Here's the basic syntax. Any specific number of arguments can be used.

```
class my_class_name {
public:
    // ... normal class stuff ...
    my_return_type operator() ( /* my list of args */ );
};
```

23.3 Why are Functors Useful?

- One example is the default 3rd argument for `std::sort`. We know that by default STL's sort routines will use the less than comparison function for the type stored inside the container. How exactly do they do that?
- First let's define another tiny helper function:

```
bool float_less(float x, float y) {
    return x < y;
}
```

- Remember how we can sort the `my_data` vector defined above using our own homemade comparison function for sorting:

```
std::sort(my_data.begin(), my_data.end(), float_less);
```

If we don't specify a 3rd argument:

```
std::sort(my_data.begin(), my_data.end());
```

This is what STL does by default:

```
std::sort(my_data.begin(), my_data.end(), std::less<float>());
```

- What is `std::less`? It's a templated class. Above we have called the default constructor to make an instance of that class. Then, that instance/object can be used like it's a function. Weird!
- How does it do that? `std::less` is a teeny tiny class that just contains the overloaded function call operator.

```
template <class T>
class less {
public:
    bool operator() (const T& x, const T& y) const { return x < y; }
};
```

You can use this instance/object/functor as a function that expects exactly two arguments of type `T` (in this example `float`) that returns a `bool`. That's exactly what we need for `std::sort`! This ultimately does the same thing as our tiny helper homemade compare function!

23.4 Another more Complicated Functor Example

- Constructors of function objects can be used to specify *internal data* for the functor that can then be used during computation of the function call operator! For example:

```
class between_values {
private:
    float low, high;
public:
    between_values(float l, float h) : low(l), high(h) {}
    bool operator() (float val) { return low <= val && val <= high; }
};
```

- The range between `low` & `high` is specified when a functor/an instance of this class is created. We might have multiple different instances of the `between_values` functor, each with their own range. Later, when the functor is used, the query value will be passed in as an argument. The function call operator accepts that single argument `val` and compares against the internal data `low` & `high`.
- This can be used in combination with STL's `find_if` construct. For example:

```
between_values two_and_four(2,4);

if (std::find_if(my_data.begin(), my_data.end(), two_and_four) != my_data.end()) {
    std::cout << "Found a value greater than 2 & less than 4!" << std::endl;
}
```

- Alternatively, we could create the functor without giving it a variable name. And in the use below we also capture the return value to print out the first item in the vector inside this range. Note that it does not print all values in the range.

```
std::vector<float>::iterator itr;
itr = std::find_if(my_data.begin(), my_data.end(), between_values(2,4));
if (itr != my_data.end()) {
    std::cout << "my_data contains " << *itr
        << ", a value greater than 2 & less than 4!" << std::endl;
}
```

23.5 Hash Table in STL?

- The Standard Template Library standard and implementation of hash table have been slowly evolving over many years. Unfortunately, the names “hashset” and “hashmap” were spoiled by developers anticipating the STL standard, so to avoid breaking or having name clashes with code using these early implementations...
- STL's agreed-upon standard for hash tables: `unordered_set` and `unordered_map`
- Depending on your OS/compiler, you may need to add the `-std=c++11` flag to the compile line (or other configuration tweaks) to access these more recent pieces of STL. (And this will certainly continue to evolve in future years!) Also, for many types STL has a good default hash function, so you may not always need to specify both template parameters!

23.6 Writing our own Hash Functions or Hash Functors

- Often the programmer/designer for the program using a hash function has the best understanding of the distribution of data to be stored in the hash function. Thus, they are in the best position to define a custom hash function (if needed) for the data & application.
- Here's an example of a (generically) good hash function for STL strings:
Note: This implementation comes from <http://www.partow.net/programming/hashfunctions/>

```
unsigned int MyHashFunction(std::string const& key) {
    unsigned int hash = 1315423911;
    for(unsigned int i = 0; i < key.length(); i++)
        hash ^= ((hash << 5) + key[i] + (hash >> 2));
    return hash;
}
```

- A functor is just a class wrapper around a function, and the function is implemented as the overloaded function call operator for the class.

```
class MyHashFunctor {
public:
    unsigned int operator() (std::string const& key) const {
        unsigned int hash = 1315423911;
        for(unsigned int i = 0; i < key.length(); i++)
            hash ^= ((hash << 5) + key[i] + (hash >> 2));
        return hash;
    }
};
```

23.7 Using STL's Associative Hash Table (Map)

- Using the default `std::string` hash function.
 - With no specified initial table size.
`std::unordered_map<std::string, Foo> m;`
 - Optionally specifying initial (minimum) table size.
`std::unordered_map<std::string, Foo> m(1000);`
- Using a home-made `std::string` hash function. Note: We are required to specify the initial table size.
 - Manually specifying the hash function type.
`std::unordered_map<std::string, Foo, std::function<unsigned int(std::string)>> m(1000, MyHashFunction);`
 - Using the `decltype` specifier to get the “declared type of an entity”.
`std::unordered_map<std::string, Foo, decltype(&MyHashFunction)> m(1000, MyHashFunction);`
- Using a home-made `std::string` hash functor or function object.
 - With no specified initial table size.
`std::unordered_map<std::string, Foo, MyHashFunctor> m;`
 - Optionally specifying initial (minimum) table size.
`std::unordered_map<std::string, Foo, MyHashFunctor> m(1000);`
- Note: In the above examples we're creating a association between two types (STL strings and custom `Foo` object). If you'd like to just create a set (no associated 2nd type), simply switch from `unordered_map` to `unordered_set` and remove the `Foo` from the template type in the examples above.

Our Copycat Version: A Set As a Hash Table, using a Hash Functor

Finish discussing Lecture 22 notes...

23.8 Homework 9 Image Comparison Hashing Discussion

- What exactly are we hashing? Do we need to write our own hash function?
- What is a good hash function for this application? What do we do when there are collisions?
- Are we allowed to use STL's hash tables (`unordered_map` and/or `unordered_set`)? Should we?
- Do we need to implement our own templated hash table class (similar to or extending `ds_hash_set`)?

