C++ONLINE

ARNO SCHOEDL

WHY ITERATORS GOT IT ALL WRONG -AND WHAT WE SHOULD USE INSTEAD

Iterators



- part of C++ since the stone age
- modeled after pointer

Iterators



- part of C++ since the stone age
- modeled after pointer
- iterators can be elements v

Iterators



- part of C++ since the stone age
- modeled after pointer
- iterators can be elements v

iterators can be borders between elements

Ranges



• anything that has iterators

```
for( auto it=ranges::begin(rng); it!=ranges::end(rng); ++it ){...}
```

Ranges



• anything that has iterators

```
for( auto it=ranges::begin(rng); it!=ranges::end(rng); ++it ){...}
```

• containers

```
vector
list
set
```

- o wn elements
- deep copying
 - copying copies elements in O(N)
- deep constness
 - const objects implies const elements

Ranges



• anything that has iterators

```
for( auto it=ranges::begin(rng); it!=ranges::end(rng); ++it ){...}
```

• containers

```
vector
list
set
```

- o wn elements
- deep copying
 - copying copies elements in O(N)
- deep constness
 - const objects implies const elements
- views

Views



- reference elements
- shallow copying
 - copying copies reference in O(1)
- shallow constness
 - view object const independent of element const

Views



- reference elements
- shallow copying
 - copying copies reference in O(1)
- shallow constness
 - view object const independent of element const

```
template < typename It >
struct subrange {
    It m_itBegin;
    It m_itEnd;
    It begin() const {
        return m_itBegin;
    }
    It end() const {
        return m_itEnd;
    }
};
```

Views (2)



- more compact code
 - with iterators:

```
std::vector<T> vec=...;
std::sort( vec.begin(), vec.end() );
vec.erase( std::unique( vec.begin, vec.end() ), vec.end() );
```

• with ranges:

```
ranges::sort(vec);
vec.erase(ranges::unique(vec).begin(),vec.end());
```

Views (2)



- more compact code
 - with iterators:

```
std::vector<T> vec=...;
std::sort( vec.begin(), vec.end() );
vec.erase( std::unique( vec.begin, vec.end() ), vec.end() );
```

with ranges:

```
ranges::sort(vec);
vec.erase(ranges::unique(vec).begin(),vec.end());
```

with think-cell ranges:

```
tc::sort_unique_inplace(vec);
```

Views (3): Range Adaptors



Views (3): Range Adaptors



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto it=ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
); // first element of value 0 in first
```

- related in semantics
- not at all related in syntax
- projection and search criterion lumped together

Transform Adaptor



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto it=ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
); // first element of value 0 in first
```

Transform Adaptor



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto it=ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
); // first element of value 0 in first
```

separation of projection and search criterion

```
auto trans=vec | views::transform(&std::pair<int,char>::first); // {0,0,1,1}
auto it=ranges::find(trans,0); // first element of value 0 in first
```

- vec not modified
- trans referencing vec
- transformation lazy
 - only pay for what you dereference
 - no extra heap memory

Transform Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
)->second; // 'a' !
```

Transform Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
)->second; // 'a' !
```

```
auto trans=vec | views::transform(&std::pair<int,char>::first); // {0,0,1,1}
ranges::find(trans,0)->second // 'a' ?
```

Transform Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
ranges::find_if(
    v,
    [](A const& a){ return a.first==0; }
)->second; // 'a' !
```

```
auto trans=vec | views::transform(&std::pair<int,char>::first); // {0,0,1,1}
ranges::find(trans,0)->second // 'a' ?
```

- iterator points to int
- peel off transform to get iterator pointing to std::pair<int,char>

Transform Adaptor (3)



- ranges::find returns iterator in role of element
- .base() must preserve identity of element

```
ranges::find(trans,0).base()->second // 'a'

v
{ 0,      0,      1,      1      } // trans
v // .base()
{{0,'a'},{0,'b'},{1,'a'},{1,'b'}} // vec
```

Transform Adaptor (3)



- ranges::find returns iterator in role of element
- .base() must preserve identity of element

- ranges::upper bound returns iterator in role of border
- same base() preserves identity of border

Filter Adaptor



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
    // {{0,'b'},{1,'b'}}
```

Filter Adaptor



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
   // {{0, 'b'},{1, 'b'}}
auto trans=filt | views::transform(&std::pair<int,char>::first); // {0,1}
ranges::find(trans,0).base().base()
      0 , 1 } // trans
      v // .base()
     {0,'b'}, {1,'b'}} // filt
         v // .base()
{{0, 'a'},{0, 'b'},{1, 'a'},{1, 'b'}} // vec
```

- ranges::find returns iterator in role of element
- result would be different without views::filter
- OK?

Filter Adaptor



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
   // {{0,'b'},{1,'b'}}
auto trans=filt | views::transform(&std::pair<int,char>::first); // {0,1}
ranges::find(trans,0).base().base()
      0 , 1 } // trans
      v // .base()
     {0,'b'}, {1,'b'}} // filt
          v // .base()
{{0, 'a'},{0, 'b'},{1, 'a'},{1, 'b'}} // vec
```

- ranges::find returns iterator in role of element
- irrelevant: result would be different without views::filter
- important: .base() preserves identity of element

Filter Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
   // {{0, 'b'},{1, 'b'}}
auto trans=filt | views::transform(&std::pair<int,char>::first); // {0,1}
ranges::upper bound(trans,0).base().base()
     {{0,'a'},{0,'b'},{1,'a'},{1,'b'}} // vec
```

- .base() preserves identity of element
- OK?

Filter Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
   // {{0,'b'},{1,'b'}}
auto trans=filt | views::transform(&std::pair<int,char>::first); // {0,1}
ranges::upper bound(trans,0).base().base()
     {{0, 'a'},{0, 'b'},{1, 'a'},{1, 'b'}} // vec
```

- .base() preserves identity of element
- identity of border not preservable
- filter(...).base() ambiguous if iterator in role of border

Filter Adaptor (2)



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto filt=vec | views::filter([](auto const& p){return p.second=='b';});
   // {{0,'b'},{1,'b'}}
auto trans=filt | views::transform(&std::pair<int,char>::first); // {0,1}
ranges::upper bound(trans,0).base().base()
     {{0, 'a'},{0, 'b'},{1, 'a'},{1, 'b'}} // vec
```

- .base() preserves identity of element
- identity of border not preservable
- filter(...).base() ambiguous if iterator in role of border
- THEN DON'T CALL IT, DUMBA**!!!



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto rev=vec | views::reverse; // {{1,'b'},{1,'a'},{0,'b'},{0,'a'}}
```



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto rev=vec | views::reverse; // {{1, 'b'},{1, 'a'},{0, 'b'},{0, 'a'}}
auto trans=rev | views::transform(&std::pair<int,char>::first); // {1,1,0,0}
ranges::find(trans,0).base().base()
{ 1 , 1 , 0 , 0 } // trans
                 v // .base()
{{1,'b'},{1,'a'},{0,'b'},{0,'a'}} // rev
                 v // .base()
       {{0,'a'},{0,'b'},{1,'a'},{1,'b'}}; // vec
```

• .base() must preserve identity of element



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto rev=vec | views::reverse; // {{1, 'b'},{1, 'a'},{0, 'b'},{0, 'a'}}
auto trans=rev | views::transform(&std::pair<int,char>::first); // {1,1,0,0}
ranges::lower bound(trans,0,std::greater<>()).base().base()
{ 1 , 1 , 0 , 0 } // trans
                 v // .base()
{{1,'b'},{1,'a'},{0,'b'},{0,'a'}} // rev
                 v // .base()
       {{0,'a'},{0,'b'},{1,'a'},{1,'b'}}; // vec
```

• .base() must preserve identity of element



```
std::vector<std::pair<int,char>> vec={{0,'a'},{0,'b'},{1,'a'},{1,'b'}};
auto rev=vec | views::reverse; // {{1, 'b'},{1, 'a'},{0, 'b'},{0, 'a'}}
auto trans=rev | views::transform(&std::pair<int,char>::first); // {1,1,0,0}
ranges::lower bound(trans,0,std::greater<>()).base().base()
{ 1 , 1 , 0 , 0 } // trans | v // .base()
{{1, 'b'},{1, 'a'},{0, 'b'},{0, 'a'}} // rev
                  {{0,'a'},{0,'b'},{1,'a'},{1,'b'}}; // vec
```

- .base() must preserve identity of element
- base() must also preserve identity of border

!!! the same base() cannot do both !!!



```
struct reverse_adaptor {
    struct iterator {
        BaseIt m_it;
        operator++() { --m_it; }
        operator--() { ++m_it; }
        operator*() { return *(m_it-1); } // why?
    };
    begin() { return iterator{m_base.end()}; }
    end() { return iterator{m_base.begin()}; }
};
```



```
struct reverse_adaptor {
    struct iterator {
        BaseIt m_it;
        operator++() { --m_it; }
        operator--() { ++m_it; }
        operator*() { return *(m_it-1); } // why?
    };
    begin() { return iterator{m_base.end()}; }
    end() { return iterator{m_base.begin()}; }
};
```

iterator at begin() stores m_it=m_base.end()
 may be dereferenced
 must return *(m base.end()-1)

• iterator at end() stores m it=m base.begin()

won't be dereferenced



```
struct reverse adaptor {
   struct iterator {
     BaseIt m it;
    operator++() { --m it; }
    operator--() { ++m it; }
    operator*() { return *(m_it-1); } // why?
  };
   begin() { return iterator{m base.end()}; }
   end() { return iterator{m base.begin()}; }
};
• iterator at begin() stores m it=m base.end()

    may be dereferenced

 o must return *(m base.end()-1)
```



```
struct reverse adaptor {
   struct iterator {
     BaseIt m it;
     operator++() { --m it; }
     operator--() { ++m it; }
     operator*() { return *(m_it-1); } // why?
   };
   begin() { return iterator{m base.end()}; }
   end() { return iterator{m base.begin()}; }
};
• iterator at begin() stores m it=m base.end()

    may be dereferenced

 o must return *(m base.end()-1)
iterator at end() stores m it=m base.begin()

    won't be dereferenced

    decrementing iterator to end()-1 makes m_it=m_base.begin()+1

 o returns *(m base.begin()+1-1) - OK!
```



```
struct reverse_adaptor {
    struct iterator {
        BaseIt m_it;
        operator++() { --m_it; }
        operator--() { ++m_it; }
        operator*() { return *(m_it-1); }
        base() { return m_it-1; } // correct for element
        base() { return m_it; } // correct for border
    };
    begin() { return iterator{m_base.end()}; }
    end() { return iterator{m_base.begin()}; }
};
```

• element *after* border in **reverse** sequence is element *before* border in original sequence

Scope of Problem: Order



• adaptor changes order of elements

Scope of Problem: Order



- adaptor changes order of elements
 - base() of element well-defined
 - base() of border in general undefined
 - example: sort adaptor

Scope of Problem: Order



- adaptor changes order of elements
 - base() of element well-defined
 - base() of border in general undefined
 - example: sort adaptor
- reverse adaptor
 - special: everything changes sides
 - base() of border well-defined, but different from base() of element

Scope of Problem: Removal



```
| v
{b, d}
|???|v//.base()
{a,b,c,d}
```

• adaptor removes elements

Scope of Problem: Removal



```
| v
{b, d}
|???|v//.base()
{a,b,c,d}
```

- adaptor removes elements
 - elements may collapse into border
 - base() of element well-defined
 - base() of border ambiguous
 - ex.: filter, sorted intersection, sorted difference

Scope of Problem: Adding



• adaptor adds elements

Scope of Problem: Adding



- adaptor adds elements
 - elements appear that were not present in base
 - base() of border well-defined
 - base() of element in general undefined
 - ex.: sorted_union

What do we do?



- Separate functions border_base and element_base
 - (+) small change
 - ∘ (-) no safety against wrong choice

What do we do?



- Separate functions border_base and element_base
 - (+) small change
 - (-) no safety against wrong choice
- Separate concepts Border and Element
 - (-) big change: no more iterator (at least in user code)
 - (+) base() always does right thing

What do we do?



- Separate functions border_base and element_base
 - (+) small change
 - (-) no safety against wrong choice
- Separate concepts Border and Element
 - (-) big change: no more iterator (at least in user code)
 - (+) base() always does right thing

Q: Do Iterators really have assigned roles Border/Element?

Hypothesis tested against our codebase (~1M LOC)

Border vs Element



- Hypothesis tested against our codebase (~1M LOC)
- find
 - 201 single match
 - 1 border role
 - 1 incremented to get border after
 - o others element role
 - 98 first match
 - o 7 border role
 - 5 incremented to get border after
 - o others element role

Border vs Element (2)



- Hypothesis tested against our codebase (~1M LOC)
- find_if
 - o 67 single match
 - o all element role
 - o 75 first match
 - o 3 border role
 - o others element role

Border vs Element (3)



- Hypothesis tested against our codebase (~1M LOC)
- lower_bound
 - 2 no further use of predicate
 - o border role
 - 89 use predicate to find single match
 - o all element role
 - 19 use predicate to find first match
 - o all element role

Border vs Element (3)



- Hypothesis tested against our codebase (~1M LOC)
- lower_bound
 - o 2 no further use of predicate
 - border role
 - 89 use predicate to find single match
 - all element role
 - 19 use predicate to find first match
 - all element role
- upper_bound
 - 24 total
 - 17 border role
 - 7 decremented to get element before

Border vs Element (3)



- Hypothesis tested against our codebase (~1M LOC)
- lower_bound
 - 2 no further use of predicate
 - o border role
 - 89 use predicate to find single match
 - all element role
 - 19 use predicate to find first match
 - o all element role
- upper_bound
 - 24 total
 - 17 border role
 - 7 decremented to get element before
- -> Iterator instances have distinct roles Border/Element

Iterators were always ugly



```
    begin() and end() asymmetric
    can dereference begin()
    cannot dereference end()
```

```
begin end
  v v v v v
{ a , b , c , d }
```

Iterators were always ugly



- begin() and end() asymmetric
 - can dereference begin()
 - cannot dereference end()

```
begin end
V V V V
{a,b,c,d}
```

• elements and borders are symmetric

```
begin end
| v | v | v | v |
{ a , b , c , d }
```

Iterators were always ugly (2)



```
auto it=ranges::find(rng, t);
if(it!=ranges::end(rng)) {...}
```

- end() 's meaning depends on role
 - if border, border after all elements
 - ∘ if element, magic value to say "none"

Iterators were always ugly (2)



```
auto it=ranges::find(rng, t);
if(it!=ranges::end(rng)) {...}
```

- end() 's meaning depends on role
 - o if border, border after all elements
 - if element, magic value to say "none"
- have to mention rng twice
 - cannot write range expression inline

Iterators were always ugly (2)



```
auto it=ranges::find(rng, t);
if(it!=ranges::end(rng)) {...}
```

- end() 's meaning depends on role
 - o if border, border after all elements
 - if element, magic value to say "none"
- have to mention rng twice
 - cannot write range expression inline
- why not

```
if( auto it=tc::find(rng, t) ) {...}
```

-> Introduce Border and Element concepts!



- Border | : like Iterator but
 - cannot be dereferenced
 - if not at begin, has element_before()
 - o if not at end, has element_after()



```
begin end
| v | v | v |
{ a , b , c , d }
```

- Border | : like Iterator but
 - cannot be dereferenced
 - if not at begin, has element_before()
 - o if not at end, has element_after()
- range begin() and end() are borders



```
begin end
| v | v | v | v |
{ a , b , c , d }
```

- Border | : like Iterator but
 - cannot be dereferenced
 - if not at begin, has element_before()
 - if not at end, has element_after()
- range begin() and end() are borders
- all iterators going into <algorithm> are borders
 - begin or end of input range



```
begin end
| v | v | v | v |
{ a , b , c , d }
```

- Border | : like Iterator but
 - cannot be dereferenced
 - if not at begin, has element_before()
 - if not at end, has element_after()
- range begin() and end() are borders
- all iterators going into <algorithm> are borders
 - begin or end of input range
- all output iterators are borders
 - begin or end of output



```
begin end | v | v | v | v | { a , b , c , d }
```

- Border | : like Iterator but
 - cannot be dereferenced
 - if not at begin, has element_before()
 - if not at end, has element_after()
- range begin() and end() are borders
- all iterators going into <algorithm> are borders
 - begin or end of input range
- all output iterators are borders
 - begin or end of output
- iterators returned from <algorithm>
 - depends on algorithm

Border Concept (2)



returned iterators are borders:

Element Concept



```
begin end
| v | v | v | v |
{ a , b , c , d }
```

- Element v : like Iterator but
 - never end(), cannot ++ beyond last element
 - o has border_before() / border_after()
- following algorithms return element:

```
find[_if]
[max_|min_]element // max/min element of a range
```

range_of_elements utility to get all elements inside borders

```
tc::for_each( range_of_elements(range), [&]( auto element ){...} );
```



- make Element nullable
 - compatible with pointer: pointer satisfies Element concept
 - contextually convertible to bool
 - null state reached through value initialization Element{}
 - functions returning Element return null instead of .end()

```
Element elem{}; // null element
assert(!elem);
if( auto it=tc::find(rng, t) ) {...}
```



• let programmer encode her intent



- let programmer encode her intent
- ranges::find[if] gets refined to

```
tc::find_unique[_if] -> Element
tc::find_first[_if] -> Element
tc::find_last[_if] -> Element
tc::trim_left[_if] -> Border
tc::trim_right[_if] -> Border
```



- let programmer encode her intent
- ranges::find[_if] gets refined to

```
tc::find_unique[_if] -> Element
tc::find_first[_if] -> Element
tc::find_last[_if] -> Element
tc::trim_left[_if] -> Border
tc::trim_right[_if] -> Border
```

• std::lower bound gets refined to

```
tc::binary_find_unique -> Element
tc::binary_find_first -> Element
tc::binary_find_last -> Element
tc::lower_bound -> Border
```



- let programmer encode her intent
- ranges::find[_if] gets refined to

```
tc::find_unique[_if] -> Element
tc::find_first[_if] -> Element
tc::find_last[_if] -> Element
tc::trim_left[_if] -> Border
tc::trim_right[_if] -> Border
```

• std::lower bound gets refined to

```
tc::binary_find_unique -> Element
tc::binary_find_first -> Element
tc::binary_find_last -> Element
tc::lower_bound -> Border
```

• can always use border_before() / border_after() to convert Element to Border



- let programmer encode her intent
- ranges::find[_if] gets refined to

```
tc::find_unique[_if] -> Element
tc::find_first[_if] -> Element
tc::find_last[_if] -> Element
tc::trim_left[_if] -> Border
tc::trim_right[_if] -> Border
```

• std::lower bound gets refined to

```
tc::binary_find_unique -> Element
tc::binary_find_first -> Element
tc::binary_find_last -> Element
tc::lower_bound -> Border
```

- can always use border_before() / border_after() to convert Element to Border
- <u>unique</u> functions assert single match

Algorithm Return Specification



```
if( auto it=tc::find_unique(rng, t) ) {...}
```

- want to mention rng only once
 - can write range expression inline

Algorithm Return Specification



```
if( auto it=tc::find_unique(rng, t) ) {...}
```

- want to mention rng only once
 - can write range expression inline
- let programmer write intent
 - algorithms get template parameter to control return value

Algorithm Return Specification



```
if( auto it=tc::find_unique(rng, t) ) {...}
```

- want to mention rng only once
 - can write range expression inline
- let programmer write intent
 - algorithms get template parameter to control return value
- algorithm returning border

```
// return border
tc::lower_bound<tc::return_border> -> border

// return view beginning/ending at border
tc::lower_bound<tc::return_take> -> view
tc::lower_bound<tc::return_drop> -> view

// return border's adjacent element
tc::lower_bound<tc::return_element_after> -> element
tc::upper_bound<tc::return_element_before> -> element
```

Algorithm Return Spec. (2)



algorithm returning element

```
// return element, which may not be there
tc::find first<tc::return element or null>
// return element, which must be there
tc::find first<tc::return element>
// return element's adjacent border [or alternative if not there]
tc::find first<tc::return border before[ or begin| or end]>
tc::find first<tc::return border after[ or begin| or end]>
// return view beginning/ending adjacent to element
// [or alternative if not there]
tc::find first<tc::return take before[ or empty| or all]>
tc::find first<tc::return take after[ or empty| or all]>
tc::find first<tc::return drop before[ or empty| or all]>
tc::find first<tc::return drop after[ or empty| or all]>
```

Return Spec. Implementation



```
template< typename Rng >
struct return_take_before_or_empty {
  template<typename It>
  static auto pack_element(It it, Rng&& rng) noexcept {
    return tc::take(std::forward<Rng>(rng), it);
  }
  static auto pack_no_element(Rng&& rng) noexcept {
    return tc::take(std::forward<Rng>(rng), ranges::begin(rng));
  }
};
```

Conclusion



- Iterator modeled after pointers
 - low level machine concept
- Element and Border stronger semantics
 - intent already in programmer's head
 - express intent in code
 - needed for correctness of important range functions
- think-cell range library https://github.com/think-cell/think-cell-library
 - Element nullable
 - algorithm refinements
 - return specifications
- still missing
 - Border not dereferencable
 - no implicit conversion Element->Border

THANK YOU!



for attending.

And yes, we are recruiting:

hr@think-cell.com