Writing Regexps 2021-22 / Regex Parser

Generated by Doxygen 1.9.3

Chapter 1

Namespace Index

1.1 Namespace List

Here is a list of all namespaces with brief descriptions:

Chapter 2

Hierarchical Index

2.1 Class Hierarchy

This inheritance list is sorted roughly, but not completely, alphabetically:

Chapter 3

Class Index

3.1 Class List

Here are the classes, structs, unions and interfaces with brief descriptions:

Chapter 4

File Index

4.1 File List

Here is a list of all files with brief descriptions:

Chapter 5

Namespace Documentation

5.1 wr22 Namespace Reference

Namespaces

- namespace regex_parser

5.2 wr22::regex_parser Namespace Reference

Namespaces

- namespace parser
- namespace regex
- namespace utils

5.3 wr22::regex_parser::parser Namespace Reference

Namespaces

- namespace errors

Classes

- class Parser

  A regex parser.

Functions

- template<typename Iter , typename Sentinel >
  Parser (Iter begin, Sentinel end) -> Parser< Iter, Sentinel >

  The type deduction guideline for `Parser`.
- regex::Part parse_regex (const utils::UnicodeStringView &regex)

  Parse a regular expression into its AST.

5.3.1 Function Documentation

5.3.1.1 parse_regex()

```
regex::Part wr22::regex_parser::parser::parse_regex (
            const utils::UnicodeStringView & regex )
```

Parse a regular expression into its AST.

The regular expression is a string view in the UTF-8 encoding (the `utils::UnicodeStringView` wrapper is used for convenience and type safety; see its docs for additional information). It is parsed and its object representation (see the docs for `regex::Part`) is built. The returned representation is an owned object and its lifetime does not depend on the lifetime of the `regex` argument.

If the parsing fails, an exception is thrown. `errors::ParseError` is the base class for all exceptions thrown from this function, but more specific exceptions may be caught and handled separately. See the docs for the `errors.hpp` file for details.

Returns

the parsed regex AST if the parsing succeeds.

Exceptions

errors::ParseError	if the parsing fails.

5.3.1.2 Parser()

```
template<typename Iter , typename Sentinel >
wr22::regex_parser::parser::Parser (
            Iter begin,
            Sentinel end ) -> Parser< Iter, Sentinel >
```

The type deduction guideline for `Parser`.

5.4 wr22::regex_parser::parser::errors Namespace Reference

Classes

- class ExpectedEnd
  The error when the parser expected the input to end, but it did not.
- struct ParseError
  The base class for parse errors.
- class UnexpectedChar
  The error when the parser got a character it didn't expect at the current position.
- class UnexpectedEnd
  The error when the parser hit the end of the input earlier than it expected.

5.5 wr22::regex_parser::regex Namespace Reference

Namespaces

- namespace capture
- namespace part

  The namespace with the variants of `Part`.

Classes

- class Capture

  Group capture behavior.
- class Part

  A part of a regular expression and its AST node type.

Enumerations

- enum class NamedCaptureFlavor { Apostrophes , Angles , AnglesWithP }

  The flavor (dialect) of a named group capture.

Functions

- std::ostream & operator<< (std::ostream &out, const Capture &capture)
- std::ostream & operator<< (std::ostream &out, NamedCaptureFlavor flavor)
- std::ostream & operator<< (std::ostream &out, const Part &part)

  Convert a `Part` to a textual representation and write it to an `std::ostream`.

5.5.1 Enumeration Type Documentation

5.5.1.1 NamedCaptureFlavor

`enum class wr22::regex_parser::regex::NamedCaptureFlavor [strong]`

The flavor (dialect) of a named group capture.

The most common variants are included. This list is subject to extension if deemed necessary. The source used as a reference is `https://www.regular-expressions.info/named.html`.

Enumerator

Apostrophes	The flavor `(?'name'contents)`. Mostly used in C# and other .NET-oriented languages, although can also be found in certain versions Perl, Boost and elsewhere.
Angles	The flavor `(?<name>contents)`. Mostly used in C# and other .NET-oriented languages, although can also be found in certain versions Perl, Boost and elsewhere.
AnglesWithP	The flavor `(?P<name>contents)`. Found in Python, PCRE and elsewhere.

5.5.2 Function Documentation

5.5.2.1 operator<<() [1/3]

```
std::ostream & wr22::regex_parser::regex::operator<< (
            std::ostream & out,
            const Capture & capture )
```

5.5.2.2 operator<<() [2/3]

```
std::ostream & wr22::regex_parser::regex::operator<< (
            std::ostream & out,
            const Part & part )
```

Convert a `Part` to a textual representation and write it to an `std::ostream`.

5.5.2.3 operator<<() [3/3]

```
std::ostream & wr22::regex_parser::regex::operator<< (
            std::ostream & out,
            NamedCaptureFlavor flavor )
```

5.6 wr22::regex_parser::regex::capture Namespace Reference

Classes

- struct Index
- struct Name
- struct None

Typedefs

- using Adt = utils::Adt< None, Index, Name >

5.6.1 Typedef Documentation

5.6.1.1 Adt

```
using wr22::regex_parser::regex::capture::Adt = typedef utils::Adt<None, Index, Name>
```

5.7 wr22::regex_parser::regex::part Namespace Reference

The namespace with the variants of `Part`.

Classes

- struct Alternatives

  A regex part with the list of alternatives to be matched.
- struct Empty

  An empty regex part.
- struct Group

  A regex part that represents a group in parentheses.
- struct Literal

  An regex part that matches a single character literally.
- struct Sequence

  A regex part with the list of items to be matched one after another.

Typedefs

- using Adt = utils::Adt< Empty, Literal, Alternatives, Sequence, Group >

5.7.1 Detailed Description

The namespace with the variants of `Part`.

See the docs for the `Part` type for additional information.

5.7.2 Typedef Documentation

5.7.2.1 Adt

```
using wr22::regex_parser::regex::part::Adt = typedef utils::Adt<Empty, Literal, Alternatives,
Sequence, Group>
```

5.8 wr22::regex_parser::utils Namespace Reference

Namespaces

- namespace detail

Classes

- class Adt

  A helper class that simplifies creation of algebraic data types.
- class UnicodeStringView

  A wrapper around `std::string_view` that enables UTF-8 codepoint iteration.
- class UnicodeStringViewIterator

  Iterator over code points for `UnicodeStringView`.

Typedefs

- using utf_traits = boost::locale::utf::utf_traits< char >

Functions

- template<typename... Variants>
  bool operator== (const Adt< Variants... > &lhs, const Adt< Variants... > &rhs)

  Compare two compatible ADTs for equality.
- template<typename... Variants>
  bool operator!= (const Adt< Variants... > &lhs, const Adt< Variants... > &rhs)

  Compare two compatible ADTs for non-equality.

5.8.1 Typedef Documentation

5.8.1.1 utf_traits

```
using wr22::regex_parser::utils::utf_traits = typedef boost::locale::utf::utf_traits<char>
```

5.8.2 Function Documentation

5.8.2.1 operator"!=()

```
template<typename... Variants>
bool wr22::regex_parser::utils::operator!= (
            const Adt< Variants... > & lhs,
            const Adt< Variants... > & rhs )
```

Compare two compatible ADTs for non-equality.

5.8.2.2 operator==()

```
template<typename...  Variants>
bool wr22::regex_parser::utils::operator== (
            const Adt< Variants...  > & lhs,
            const Adt< Variants...  > & rhs )
```

Compare two compatible ADTs for equality.

5.9 wr22::regex_parser::utils::detail Namespace Reference

Namespaces

- namespace adt

5.10 wr22::regex_parser::utils::detail::adt Namespace Reference

Classes

- struct MultiCallable

Chapter 6

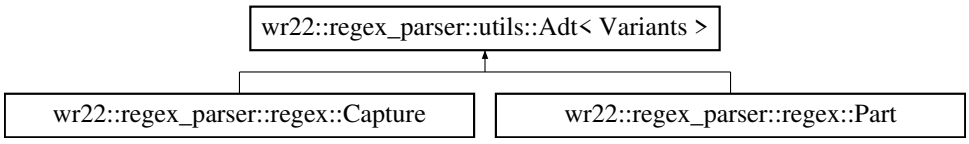
Class Documentation

6.1 wr22::regex_parser::utils::Adt< Variants > Class Template Reference

A helper class that simplifies creation of algebraic data types.

`#include <adt.hpp>`

Inheritance diagram for wr22::regex_parser::utils::Adt< Variants >:



Public Types

- using VariantType = std::variant< Variants... >

  A convenience type alias for the concrete `std::variant` type used.

Public Member Functions

- template<typename V >
  Adt (V variant)

  Constructor for each of the variants.
- template<typename... Fs>
  decltype(auto) visit (Fs &&... visitors) const

  Visit the ADT, applying the suitable function from the list of visitors on the variant held.
- template<typename... Fs>
  decltype(auto) visit (Fs &&... visitors)

  Visit the ADT, applying the suitable function from the list of visitors on the variant held.
- const VariantType & as_variant () const

  Access the underlying `std::variant` type (constant version).
- VariantType & as_variant ()

  Access the underlying `std::variant` type (non-constant version).

Protected Attributes

- VariantType m_variant

6.1.1 Detailed Description

template<typename... Variants>
class wr22::regex_parser::utils::Adt< Variants >

A helper class that simplifies creation of algebraic data types.

Algebraic data types are data types that can have one type of a predefined set of variants, but be stored and represented as values of one common type. In C++, `std::variant` serves exactly this purpose. It is, however, not very convenient to work with or build upon, so this class is designed to simplify building new algebraic data types. It still uses `std::variant` under the hood.

The template type parameters are the types that the variants may hold (must be distinct types).

6.1.2 Member Typedef Documentation

6.1.2.1 VariantType

```
template<typename...  Variants>
using wr22::regex_parser::utils::Adt< Variants >::VariantType = std::variant<Variants...>
```

A convenience type alias for the concrete `std::variant` type used.

6.1.3 Constructor & Destructor Documentation

6.1.3.1 Adt()

```
template<typename...  Variants>
template<typename V >
wr22::regex_parser::utils::Adt< Variants >::Adt (
            V variant )  [inline]
```

Constructor for each of the variants.

Construct an instance holding a specified variant. The type `V` of the variant provided must be one of the types from `Variants`. Note that this constructor is purposefully implicit, so that the variants as separate types are transparently converted to this common type when necessary.

The variant is taken by value and moved thereafter, so that, when constructing the common type, the variant may be either copied or moved, depending on the user's intentions.

6.1.4 Member Function Documentation

6.1.4.1 as_variant() [1/2]

```
template<typename... Variants>
VariantType & wr22::regex_parser::utils::Adt< Variants >::as_variant ( ) [inline]
```

Access the underlying `std::variant` type (non-constant version).

6.1.4.2 as_variant() [2/2]

```
template<typename... Variants>
const VariantType & wr22::regex_parser::utils::Adt< Variants >::as_variant ( ) const [inline]
```

Access the underlying `std::variant` type (constant version).

6.1.4.3 visit() [1/2]

```
template<typename... Variants>
template<typename... Fs>
decltype(auto) wr22::regex_parser::utils::Adt< Variants >::visit (
            Fs &&... visitors ) [inline]
```

Visit the ADT, applying the suitable function from the list of visitors on the variant held.

This is the non-constant version of the method. See the docs for the constant version for a detailed description and code examples. The only thing different in this version of the method is that the visitors get called with a non-const lvalue reference to the variants instead of a const reference.

6.1.4.4 visit() [2/2]

```
template<typename... Variants>
template<typename... Fs>
decltype(auto) wr22::regex_parser::utils::Adt< Variants >::visit (
            Fs &&... visitors ) const [inline]
```

Visit the ADT, applying the suitable function from the list of visitors on the variant held.

Using this method is essentially the same as using `std::visit` on the variant, except that, for convenience, multiple visitors are joined into one big visitor. That is, a typical `Adt` usage might look like this:

```
struct MyAdt : public Adt<int, double> {
    // Make the constructor available in the derived class.
    using Adt<int, double>::Adt;
};
// <...>
void func() {
    // Variant type: double.
    MyAdt my_adt = 3.14;
    // Prints "Double: 3.14".
    my_adt.visit(
        [](int x)    { std::cout « "Int: "    « x « std::endl; },
        [](double x) { std::cout « "Double: " « x « std::endl; }
    );
}
```

This is the constant version of the method. Visitors must be callable with the const reference to variant types.

6.1.5 Member Data Documentation

6.1.5.1 m_variant

```
template<typename...  Variants>
VariantType wr22::regex_parser::utils::Adt< Variants >::m_variant  [protected]
```

The documentation for this class was generated from the following file:

- include/wr22/regex_parser/utils/adt.hpp

6.2 wr22::regex_parser::regex::part::Alternatives Struct Reference

A regex part with the list of alternatives to be matched.

`#include <part.hpp>`

Public Member Functions

- bool operator== (const Alternatives &rhs) const =default

Public Attributes

- std::vector< Part > alternatives
  The list of the alternatives.

6.2.1 Detailed Description

A regex part with the list of alternatives to be matched.

Alternatives in regular expressions are subexpressions by |. For the whole expression part's match to succeed, at least one of the subexpressions must match the input successfully.

As an example, `a|(b)|cde` would be represented as an `Alternatives` part with 3 alternatives. The alternatives themselves are represented recursively as `Part`s.

6.2.2 Member Function Documentation

6.2.2.1 operator==()

```
bool wr22::regex_parser::regex::part::Alternatives::operator== (
            const Alternatives & rhs ) const  [default]
```

6.2.3 Member Data Documentation

6.2.3.1 alternatives

`std::vector<Part> wr22::regex_parser::regex::part::Alternatives::alternatives`

The list of the alternatives.

The documentation for this struct was generated from the following file:

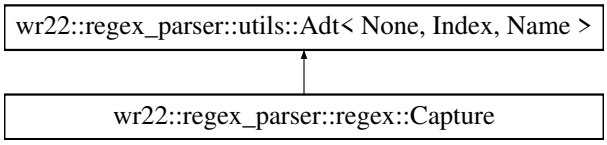
- include/wr22/regex_parser/regex/part.hpp

6.3 wr22::regex_parser::regex::Capture Class Reference

Group capture behavior.

`#include <capture.hpp>`

Inheritance diagram for wr22::regex_parser::regex::Capture:



Additional Inherited Members

6.3.1 Detailed Description

Group capture behavior.

A group can be captured by index (when one writes `(contents)`), by name (e.g. `(?<name>contents)` in some dialects) or not captured at all (`(?:contents)`). Objects of this type determine how exactly a certain group is going to be captured. This is a variant type (see `Part` and `utils::Adt` for a more detailed explanation of the concept). The variants for this class (explicitly or implicitly convertible to this type) are located in the `capture` namespace.

The documentation for this class was generated from the following file:

- include/wr22/regex_parser/regex/capture.hpp

6.4 wr22::regex_parser::regex::part::Empty Struct Reference

An empty regex part.

`#include <part.hpp>`

Public Member Functions

- bool operator== (const Empty &rhs) const =default

6.4.1 Detailed Description

An empty regex part.

Corresponds to an empty regular expression (`""`) or the contents of an empty parenthesized group (`"()"`).

6.4.2 Member Function Documentation

6.4.2.1 operator==()

```
bool wr22::regex_parser::regex::part::Empty::operator== (
            const Empty & rhs ) const  [default]
```

The documentation for this struct was generated from the following file:

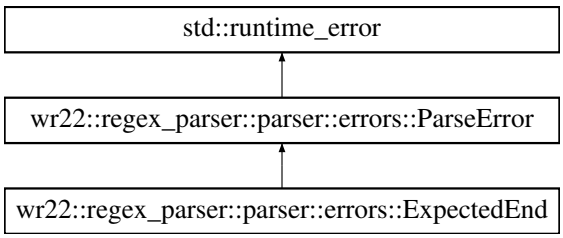
- include/wr22/regex_parser/regex/part.hpp

6.5 wr22::regex_parser::parser::errors::ExpectedEnd Class Reference

The error when the parser expected the input to end, but it did not.

`#include <errors.hpp>`

Inheritance diagram for wr22::regex_parser::parser::errors::ExpectedEnd:



Public Member Functions

- ExpectedEnd (size_t position, char32_t char_got)

  Constructor.
- size_t position () const

  Get the input position. See the constructor docs for a more detailed description.
- char32_t char_got () const

  Get the character the parser has received.

6.5.1 Detailed Description

The error when the parser expected the input to end, but it did not.

6.5.2 Constructor & Destructor Documentation

6.5.2.1 ExpectedEnd()

```
wr22::regex_parser::parser::errors::ExpectedEnd::ExpectedEnd (
            size_t position,
            char32_t char_got )
```

Constructor.

Parameters

position	the 0-based position in the input when the parser has encountered the end of input.
char_got	the character that the parser has received instead of the end of input.

6.5.3 Member Function Documentation

6.5.3.1 char_got()

```
char32_t wr22::regex_parser::parser::errors::ExpectedEnd::char_got ( ) const
```

Get the character the parser has received.

See the constructor docs for a more detailed description.

6.5.3.2 position()

```
size_t wr22::regex_parser::parser::errors::ExpectedEnd::position ( ) const
```

Get the input position. See the constructor docs for a more detailed description.

The documentation for this class was generated from the following files:

- include/wr22/regex_parser/parser/errors.hpp
- src/parser/errors.cpp

6.6 wr22::regex_parser::regex::part::Group Struct Reference

A regex part that represents a group in parentheses.

`#include <part.hpp>`

Public Member Functions

- Group (Capture capture, Part inner)
  Convenience constructor.
- bool operator== (const Group &rhs) const

Public Attributes

- Capture capture
  Capture behavior.
- std::unique_ptr< Part > inner
  The (smart) pointer to the group contents.

6.6.1 Detailed Description

A regex part that represents a group in parentheses.

A group in regular expressions is virtually everything that is enclosed with parentheses: `(some group)`, `(?:blablabla)` and `(?P<group_name>group contents)` are all groups.

A group has two main attributes: (1) how it is captured during matching and (2) the contents of the group. The contents is simply another `Part`. The capture behavior is expressed by a separate type `Capture`. See its docs for additional info, and take a look at `https://www.regular-expressions.info/brackets.html` for an introduction to or a recap of regex groups and capturing.

6.6.2 Constructor & Destructor Documentation

6.6.2.1 Group()

```
wr22::regex_parser::regex::part::Group::Group (
            Capture capture,
            Part inner )
```

Convenience constructor.

6.6.3 Member Function Documentation

6.6.3.1 operator==()

```
bool wr22::regex_parser::regex::part::Group::operator== (
            const Group & rhs ) const
```

6.6.4 Member Data Documentation

6.6.4.1 capture

```
Capture wr22::regex_parser::regex::part::Group::capture
```

Capture behavior.

6.6.4.2 inner

```
std::unique_ptr<Part> wr22::regex_parser::regex::part::Group::inner
```

The (smart) pointer to the group contents.

The documentation for this struct was generated from the following files:

- include/wr22/regex_parser/regex/part.hpp
- src/regex/part.cpp

6.7 wr22::regex_parser::regex::capture::Index Struct Reference

```
#include <capture.hpp>
```

Public Member Functions

- bool operator== (const Index &rhs) const =default

6.7.1 Member Function Documentation

6.7.1.1 operator==()

```
bool wr22::regex_parser::regex::capture::Index::operator== (
            const Index & rhs ) const  [default]
```

The documentation for this struct was generated from the following file:

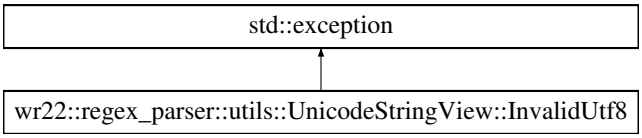
- include/wr22/regex_parser/regex/capture.hpp

6.8 wr22::regex_parser::utils::UnicodeStringView::InvalidUtf8 Struct Reference

An error thrown when the string's encoding is not valid UTF-8.

`#include <utf8_string_view.hpp>`

Inheritance diagram for wr22::regex_parser::utils::UnicodeStringView::InvalidUtf8:



Public Member Functions

- const char * what () const noexcept override

6.8.1 Detailed Description

An error thrown when the string's encoding is not valid UTF-8.

6.8.2 Member Function Documentation

6.8.2.1 what()

```
const char * wr22::regex_parser::utils::UnicodeStringView::InvalidUtf8::what ( ) const  [override],
[noexcept]
```

The documentation for this struct was generated from the following files:

- include/wr22/regex_parser/utils/utf8_string_view.hpp
- src/utils/utf8_string_view.cpp

6.9 wr22::regex_parser::regex::part::Literal Struct Reference

An regex part that matches a single character literally.

`#include <part.hpp>`

Public Member Functions

- bool operator== (const Literal &rhs) const =default

Public Attributes

- char32_t character

6.9.1 Detailed Description

An regex part that matches a single character literally.

Corresponds to a plain character in a regular expression. E.g. the regex `"foo"` contains three character literals: `f`, `o` and `o`.

6.9.2 Member Function Documentation

6.9.2.1 operator==()

```
bool wr22::regex_parser::regex::part::Literal::operator== (
            const Literal & rhs ) const  [default]
```

6.9.3 Member Data Documentation

6.9.3.1 character

```
char32_t wr22::regex_parser::regex::part::Literal::character
```

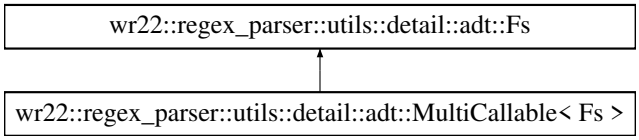
The documentation for this struct was generated from the following file:

- include/wr22/regex_parser/regex/part.hpp

6.10 wr22::regex_parser::utils::detail::adt::MultiCallable< Fs > Struct Template Reference

`#include <adt.hpp>`

Inheritance diagram for wr22::regex_parser::utils::detail::adt::MultiCallable< Fs >:



Public Member Functions

- MultiCallable (Fs &&... fs)

6.10.1 Constructor & Destructor Documentation

6.10.1.1 MultiCallable()

```
template<typename... Fs>
wr22::regex_parser::utils::detail::adt::MultiCallable< Fs >::MultiCallable (
            Fs &&... fs ) [inline]
```

The documentation for this struct was generated from the following file:

- include/wr22/regex_parser/utils/adt.hpp

6.11 wr22::regex_parser::regex::capture::Name Struct Reference

`#include <capture.hpp>`

Public Member Functions

- bool operator== (const Name &rhs) const =default

Public Attributes

- std::string name
- NamedCaptureFlavor flavor

6.11.1 Member Function Documentation

6.11.1.1 operator==()

```
bool wr22::regex_parser::regex::capture::Name::operator== (
            const Name & rhs ) const  [default]
```

6.11.2 Member Data Documentation

6.11.2.1 flavor

```
NamedCaptureFlavor wr22::regex_parser::regex::capture::Name::flavor
```

6.11.2.2 name

```
std::string wr22::regex_parser::regex::capture::Name::name
```

The documentation for this struct was generated from the following file:

- include/wr22/regex_parser/regex/capture.hpp

6.12 wr22::regex_parser::regex::capture::None Struct Reference

```
#include <capture.hpp>
```

Public Member Functions

- bool operator== (const None &rhs) const =default

6.12.1 Member Function Documentation

6.12.1.1 operator==()

```
bool wr22::regex_parser::regex::capture::None::operator== (
            const None & rhs ) const  [default]
```

The documentation for this struct was generated from the following file:

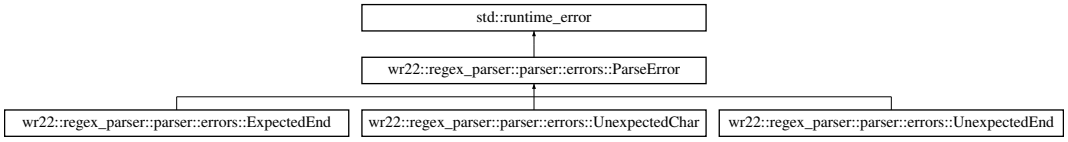
- include/wr22/regex_parser/regex/capture.hpp

6.13 wr22::regex_parser::parser::errors::ParseError Struct Reference

The base class for parse errors.

`#include <errors.hpp>`

Inheritance diagram for wr22::regex_parser::parser::errors::ParseError:



6.13.1 Detailed Description

The base class for parse errors.

This exception type should be caught if it is desired to catch all parse errors. However, there are more specific exceptions deriving from this one that can be handled separately for greater flexibility.

The documentation for this struct was generated from the following file:

- include/wr22/regex_parser/parser/errors.hpp

6.14 wr22::regex_parser::parser::Parser< Iter, Sentinel > Class Template Reference

A regex parser.

Public Member Functions

- Parser (Iter begin, Sentinel end)

  Constructor.
- void expect_end ()

  Ensure that the parser has consumed all of the input.
- regex::Part parse_regex ()

  Parse a regex consuming part of the remaining input.
- regex::Part parse_alternatives ()

  Intermediate rule: parse a pipe-separated list of alternatives (e.g.
- regex::Part parse_sequence ()

  Intermediate rule: parse a sequence of atoms (e.g.
- regex::Part parse_sequence_or_empty ()

  Intermediate rule: parse a possibly empty sequence of atoms.
- regex::Part parse_atom ()

  Intermediate rule: parse an atom.
- regex::Part parse_char_literal ()

  Intermediate rule: parse a character literal.
- regex::Part parse_group ()

  Intermediate rule: parse a parenthesized group (any capture variant).
- std::string parse_group_name ()

  Intermediate rule: parse a group name.

6.14.1 Detailed Description

template<typename Iter, typename Sentinel>
requires requires(Iter iter, Sentinel end) { ++iter; { *iter } -> std::convertible_to<char32_t>; { iter == end } -> std::convertible_to<bool>; { iter != end } -> std::convertible_to<bool>; }
class wr22::regex_parser::parser::Parser< Iter, Sentinel >

A regex parser.

For additional information see the methods' docs, particularly the constructor and the `parse_regex` method.

6.14.2 Constructor & Destructor Documentation

6.14.2.1 Parser()

```
template<typename Iter , typename Sentinel >
wr22::regex_parser::parser::Parser< Iter, Sentinel >::Parser (
            Iter begin,
            Sentinel end )  [inline]
```

Constructor.

This constructor stores a pair of forward iterators that should generate a sequence of Unicode code points (`char32_t`). The `begin` iterator and the `end` sentinel may have different types provided that the iterator can is equality comparable with the sentinel.

SAFETY: The iterators must not be invalidated as long as this `Parser` object is still alive.

6.14.3 Member Function Documentation

6.14.3.1 expect_end()

```
template<typename Iter , typename Sentinel >
void wr22::regex_parser::parser::Parser< Iter, Sentinel >::expect_end ( )  [inline]
```

Ensure that the parser has consumed all of the input.

Does nothing if all input has been consumed.

Exceptions

errors::ExpectedEnd	if this is not the case.

6.14.3.2 parse_alternatives()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_alternatives ( )
[inline]
```

Intermediate rule: parse a pipe-separated list of alternatives (e.g.

`a|bb|ccc`).

Returns

the list of parsed alternatives packed into `regex::part::Alternatives` or, if and only if the list of alternatives contains exactly 1 element, the only alternative unchanged.

Exceptions

errors::ParseError	if the input cannot be parsed.

6.14.3.3 parse_atom()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_atom ( )  [inline]
```

Intermediate rule: parse an atom.

Currently, this grammar only recognizes two kinds of atoms: character literals (individual plain characters in a regex) and parenthesized groups. As the project development goes on, new kinds of atoms will be added.

Returns

the parsed atom (some variant of `regex::Part` depending on the atom kind).

Exceptions

errors::ParseError	if the input cannot be parsed.

6.14.3.4 parse_char_literal()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_char_literal ( )
[inline]
```

Intermediate rule: parse a character literal.

Returns

the parsed character literal (`regex::part::Literal`).

Exceptions

errors::UnexpectedEnd	if all characters from the input have already been consumed.

6.14.3.5 parse_group()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_group ( )  [inline]
```

Intermediate rule: parse a parenthesized group (any capture variant).

Returns

the parsed group (`regex::part::Group`).

6.14.3.6 parse_group_name()

```
template<typename Iter , typename Sentinel >
std::string wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_group_name ( )  [inline]
```

Intermediate rule: parse a group name.

Returns

the UTF-8 encoded group name as an `std::string`.

6.14.3.7 parse_regex()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_regex ( )  [inline]
```

Parse a regex consuming part of the remaining input.

This is **the** method that should be called to parse a regular expression because it represents the root rule of the regex grammar. Please note that this method may not consume all of the parser's input. Hence, if a whole regex is to be parsed, the `expect_end` method should be called afterwards.

Returns

the parsed regex AST (some variant of `regex::Part` depending on the input).

Exceptions

errors::ParseError	if the input cannot be parsed.

6.14.3.8 parse_sequence()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_sequence ( )  [inline]
```

Intermediate rule: parse a sequence of atoms (e.g.

`a(?:b)[c-e]`).

Returns

the list of parsed atoms packed into `regex::part::Sequence` or, if and only if this list of contains exactly 1 element, the only atom unchanged.

Exceptions

errors::ParseError	if the input cannot be parsed.

6.14.3.9 parse_sequence_or_empty()

```
template<typename Iter , typename Sentinel >
regex::Part wr22::regex_parser::parser::Parser< Iter, Sentinel >::parse_sequence_or_empty ( )
[inline]
```

Intermediate rule: parse a possibly empty sequence of atoms.

Returns

`regex::part::Empty` if the sequence is empty, or calls `parse_sequence` otherwise.

Exceptions

errors::ParseError	if the input cannot be parsed.

The documentation for this class was generated from the following file:

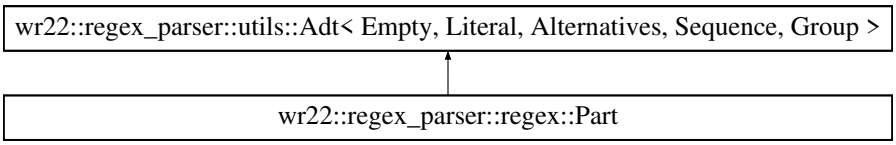
- src/parser/regex.cpp

6.15 wr22::regex_parser::regex::Part Class Reference

A part of a regular expression and its AST node type.

`#include <part.hpp>`

Inheritance diagram for wr22::regex_parser::regex::Part:



Additional Inherited Members

6.15.1 Detailed Description

A part of a regular expression and its AST node type.

The parsed regular expressions are represented as abstract syntax trees (ASTs). These are tree-like data structures where each node represents a regular expression part (or the whole regex), and, depending on their type, these nodes may have subexpressions. Subexpressions are `Part`s themselves, which also have child expressions and so on. For example, `part::Sequence` has a number of subexpressions, and each of them is of the type `Part` and is an AST node.

Each regex part has its own simple function. For example, `part::Alternatives` tries to match several alternative subexpressions against the input and succeeds if at least one of them does; and `part::Sequence` matches several subexpressions one after another, requiring them all to match respective parts of the input. By combining these simple nodes, it becomes possible to represent complex regular expressions. For example, the regex `aaa|bb` can be represented as a `part::Alternatives`, where each of the alternatives is a `parts::Sequence` of `part::Literal`s.

The `Part` itself is represented by `std::variant` via the helper class `utils::Adt`. In a nutshell, it allows a regex part to "have" one of the several predefined types (the so-called variants, which are defined in the `part` namespace), but still be represented as a `Part`. For the list of operations that can be performed on this type, e.g. to check if an instance of `Parts` has a specific variant and, if yes, access the value of this variant, see the documentation for the `utils::Adt` class, which `Part` inherits from.

The documentation for this class was generated from the following file:

- include/wr22/regex_parser/regex/part.hpp

6.16 wr22::regex_parser::regex::part::Sequence Struct Reference

A regex part with the list of items to be matched one after another.

`#include <part.hpp>`

Public Member Functions

- bool operator== (const Sequence &rhs) const =default

Public Attributes

- std::vector< Part > items

  The list of the subexpressions.

6.16.1 Detailed Description

A regex part with the list of items to be matched one after another.

Sequences in regular expressions are just subexpressions going directly one after another. As an example, `a[b-e].` is a sequence of 3 subexpressions: `a`, `[b-e]` and `.`. As an another example, `ab` is a sequence of 2 subexpressions: `a` and `b`.

6.16.2 Member Function Documentation

6.16.2.1 operator==()

```
bool wr22::regex_parser::regex::part::Sequence::operator== (
            const Sequence & rhs ) const  [default]
```

6.16.3 Member Data Documentation

6.16.3.1 items

```
std::vector<Part> wr22::regex_parser::regex::part::Sequence::items
```

The list of the subexpressions.

The documentation for this struct was generated from the following file:

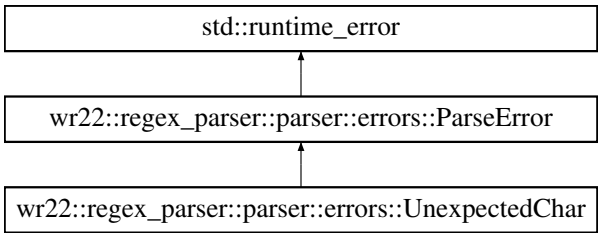
- include/wr22/regex_parser/regex/part.hpp

6.17 wr22::regex_parser::parser::errors::UnexpectedChar Class Reference

The error when the parser got a character it didn't expect at the current position.

`#include <errors.hpp>`

Inheritance diagram for wr22::regex_parser::parser::errors::UnexpectedChar:



Public Member Functions

- UnexpectedChar (size_t position, char32_t char_got, std::string expected)
  Constructor.
- size_t position () const
  Get the input position. See the constructor docs for a more detailed description.
- char32_t char_got () const
  Get the character the parser has received.
- const std::string & expected () const
  Get the description of expected characters.

6.17.1 Detailed Description

The error when the parser got a character it didn't expect at the current position.

6.17.2 Constructor & Destructor Documentation

6.17.2.1 UnexpectedChar()

```
wr22::regex_parser::parser::errors::UnexpectedChar::UnexpectedChar (
            size_t position,
            char32_t char_got,
            std::string expected )
```

Constructor.

Parameters

position	the 0-based position in the input when the parser has encountered the unexpected character.
char_got	the character that the parser has received.
expected	a textual description of a class of characters expected instead.

6.17.3 Member Function Documentation

6.17.3.1 char_got()

```
char32_t wr22::regex_parser::parser::errors::UnexpectedChar::char_got ( ) const
```

Get the character the parser has received.

See the constructor docs for a more detailed description.

6.17.3.2 expected()

```
const std::string & wr22::regex_parser::parser::errors::UnexpectedChar::expected ( ) const
```

Get the description of expected characters.

See the constructor docs for a more detailed description.

6.17.3.3 position()

```
size_t wr22::regex_parser::parser::errors::UnexpectedChar::position ( ) const
```

Get the input position. See the constructor docs for a more detailed description.

The documentation for this class was generated from the following files:

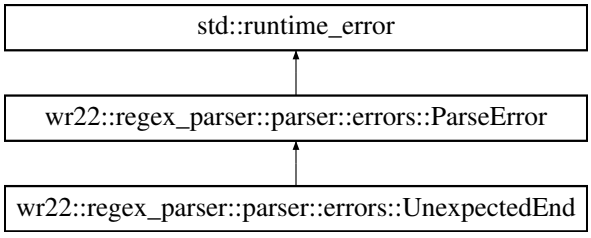
- include/wr22/regex_parser/parser/errors.hpp
- src/parser/errors.cpp

6.18 wr22::regex_parser::parser::errors::UnexpectedEnd Class Reference

The error when the parser hit the end of the input earlier than it expected.

```
#include <errors.hpp>
```

Inheritance diagram for wr22::regex_parser::parser::errors::UnexpectedEnd:



Public Member Functions

- UnexpectedEnd (size_t position, std::string expected)

  Constructor.
- size_t position () const

  Get the input position. See the constructor docs for a more detailed description.
- const std::string & expected () const

  Get the description of expected characters.

6.18.1 Detailed Description

The error when the parser hit the end of the input earlier than it expected.

6.18.2 Constructor & Destructor Documentation

6.18.2.1 UnexpectedEnd()

```
wr22::regex_parser::parser::errors::UnexpectedEnd::UnexpectedEnd (
            size_t position,
            std::string expected )
```

Constructor.

Parameters

position	the 0-based position in the input when the parser has encountered the end of input.
expected	a textual description of a class of characters expected instead.

6.18.3 Member Function Documentation

6.18.3.1 expected()

```
const std::string & wr22::regex_parser::parser::errors::UnexpectedEnd::expected ( ) const
```

Get the description of expected characters.

See the constructor docs for a more detailed description.

6.18.3.2 position()

```
size_t wr22::regex_parser::parser::errors::UnexpectedEnd::position ( ) const
```

Get the input position. See the constructor docs for a more detailed description.

The documentation for this class was generated from the following files:

- include/wr22/regex_parser/parser/errors.hpp
- src/parser/errors.cpp

6.19 wr22::regex_parser::utils::UnicodeStringView Class Reference

A wrapper around `std::string_view` that enables UTF-8 codepoint iteration.

```
#include <utf8_string_view.hpp>
```

Classes

- struct InvalidUtf8

  An error thrown when the string's encoding is not valid UTF-8.

Public Member Functions

- UnicodeStringView (const std::string_view &raw)

  Constructor.
- const std::string_view & raw () const

  Access the wrapped `std::string_view`.
- UnicodeStringViewIterator begin () const

  Create an iterator to the beginning of the string.
- UnicodeStringViewIterator end () const

  Create an iterator past the end of the string.

Friends

- class UnicodeStringViewIterator

6.19.1 Detailed Description

A wrapper around `std::string_view` that enables UTF-8 codepoint iteration.

An `UnicodeStringView` holds a "raw" `std::string_view` and assumes it is UTF-8 encoded. It provides the `begin()` & `end()` methods, which return iterators that work with Unicode codepoints instead of raw bytes. This makes this type useful in contexts where one needs to iterate over Unicode code points (like characters, but technically a different thing) in an `std::string` or an `std::string_view`.

6.19.2 Constructor & Destructor Documentation

6.19.2.1 UnicodeStringView()

```
wr22::regex_parser::utils::UnicodeStringView::UnicodeStringView (
            const std::string_view & raw )  [explicit]
```

Constructor.

Wrap an existing `std::string_view`.

The ownership of the data pointed to by `raw` is not taken, and the string contents are not copied. Hence, just like with an ordinary `std::string_view`, if the pointed data is invalidated, the `UnicodeStringView` referring to it is invalidated as well.

This constructor does not check that `raw` is a correct UTF-8 encoded string.

SAFETY: If the string contents change during the `UnicodeStringView`'s lifetime or the string is destroyed, the behavior is undefined.

6.19.3 Member Function Documentation

6.19.3.1 begin()

```
UnicodeStringViewIterator wr22::regex_parser::utils::UnicodeStringView::begin ( ) const
```

Create an iterator to the beginning of the string.

SAFETY: The returned iterator must not outlive this object.

6.19.3.2 end()

```
UnicodeStringViewIterator wr22::regex_parser::utils::UnicodeStringView::end ( ) const
```

Create an iterator past the end of the string.

SAFETY: The returned iterator must not outlive this object.

6.19.3.3 raw()

```
const std::string_view & wr22::regex_parser::utils::UnicodeStringView::raw ( ) const
```

Access the wrapped `std::string_view`.

6.19.4 Friends And Related Function Documentation

6.19.4.1 UnicodeStringViewIterator

```
friend class UnicodeStringViewIterator  [friend]
```

The documentation for this class was generated from the following files:

- include/wr22/regex_parser/utils/utf8_string_view.hpp
- src/utils/utf8_string_view.cpp

6.20 wr22::regex_parser::utils::UnicodeStringViewIterator Class Reference

Iterator over code points for `UnicodeStringView`.

```
#include <utf8_string_view.hpp>
```

Public Types

- using value_type = char32_t

  The type of value this iterator yields. Part of the STL iterator interface.
- using category_type = std::forward_iterator_tag

  This iterator's category. Part of the STL iterator interface.

Public Member Functions

- UnicodeStringViewIterator ()=delete

  The default constructor is meaningless and is thus deleted.
- UnicodeStringViewIterator & operator++ ()

  Iterator interface: pre-increment.
- char32_t operator* ()

  Iterator interface: dereferencing.
- bool operator== (const UnicodeStringViewIterator &rhs) const
- bool operator!= (const UnicodeStringViewIterator &rhs) const

Friends

- class UnicodeStringView

6.20.1 Detailed Description

Iterator over code points for `UnicodeStringView`.

SAFETY: No object of this class must outlive the `UnicodeStringView` that has created it. If this is violated, the behavior is undefined.

6.20.2 Member Typedef Documentation

6.20.2.1 category_type

```
using wr22::regex_parser::utils::UnicodeStringViewIterator::category_type = std::forward_↩
iterator_tag
```

This iterator's category. Part of the STL iterator interface.

6.20.2.2 value_type

```
using wr22::regex_parser::utils::UnicodeStringViewIterator::value_type = char32_t
```

The type of value this iterator yields. Part of the STL iterator interface.

6.20.3 Constructor & Destructor Documentation

6.20.3.1 UnicodeStringViewIterator()

```
wr22::regex_parser::utils::UnicodeStringViewIterator::UnicodeStringViewIterator ( )  [delete]
```

The default constructor is meaningless and is thus deleted.

6.20.4 Member Function Documentation

6.20.4.1 operator"!=()

```
bool wr22::regex_parser::utils::UnicodeStringViewIterator::operator!= (
            const UnicodeStringViewIterator & rhs ) const
```

6.20.4.2 operator∗()

```
char32_t wr22::regex_parser::utils::UnicodeStringViewIterator::operator* ( )
```

Iterator interface: dereferencing.

Exceptions

UnicodeStringView::InvalidUtf8	if the codepoint decodes into an invalid or incomplete value.

6.20.4.3 operator++()

```
UnicodeStringViewIterator & wr22::regex_parser::utils::UnicodeStringViewIterator::operator++ (
)
```

Iterator interface: pre-increment.

Exceptions

UnicodeStringView::InvalidUtf8	if the codepoint decodes into an invalid or incomplete value.

6.20.4.4 operator==()

```
bool wr22::regex_parser::utils::UnicodeStringViewIterator::operator== (
            const UnicodeStringViewIterator & rhs ) const
```

6.20.5 Friends And Related Function Documentation

6.20.5.1 UnicodeStringView

```
friend class UnicodeStringView  [friend]
```

The documentation for this class was generated from the following files:

- include/wr22/regex_parser/utils/utf8_string_view.hpp
- src/utils/utf8_string_view.cpp

Chapter 7

File Documentation

7.1 include/wr22/regex_parser/parser/errors.hpp File Reference

```
#include <exception>
#include <stdexcept>
#include <string>
```

Classes

- struct wr22::regex_parser::parser::errors::ParseError

  The base class for parse errors.
- class wr22::regex_parser::parser::errors::UnexpectedEnd

  The error when the parser hit the end of the input earlier than it expected.
- class wr22::regex_parser::parser::errors::ExpectedEnd

  The error when the parser expected the input to end, but it did not.
- class wr22::regex_parser::parser::errors::UnexpectedChar

  The error when the parser got a character it didn't expect at the current position.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::parser
- namespace wr22::regex_parser::parser::errors

7.2 errors.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // STL
4 #include <exception>
5 #include <stdexcept>
6 #include <string>
7
8 namespace wr22::regex_parser::parser::errors {
9
15 struct ParseError : public std::runtime_error {
16     using std::runtime_error::runtime_error;
17 };
18
20 class UnexpectedEnd : public ParseError {
21 public:
27     UnexpectedEnd(size_t position, std::string expected);
28
30     size_t position() const;
33     const std::string& expected() const;
34
35 private:
36     size_t m_position;
37     std::string m_expected;
38 };
39
41 class ExpectedEnd : public ParseError {
42 public:
48     ExpectedEnd(size_t position, char32_t char_got);
49
51     size_t position() const;
54     char32_t char_got() const;
55
56 private:
57     size_t m_position;
58     char32_t m_char_got;
59 };
60
62 class UnexpectedChar : public ParseError {
63 public:
70     UnexpectedChar(size_t position, char32_t char_got, std::string expected);
71
73     size_t position() const;
76     char32_t char_got() const;
79     const std::string& expected() const;
80
81 private:
82     size_t m_position;
83     char32_t m_char_got;
84     std::string m_expected;
85 };
86
87 }  // namespace wr22::regex_parser::parser::errors
```

7.3 include/wr22/regex_parser/parser/regex.hpp File Reference

```
#include <wr22/regex_parser/regex/part.hpp>
#include <wr22/regex_parser/utils/utf8_string_view.hpp>
#include <string_view>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::parser

Functions

- regex::Part wr22::regex_parser::parser::parse_regex (const utils::UnicodeStringView &regex)

  Parse a regular expression into its AST.

7.4 regex.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // wr22
4 #include <wr22/regex_parser/regex/part.hpp>
5 #include <wr22/regex_parser/utils/utf8_string_view.hpp>
6
7 // stl
8 #include <string_view>
9
10 namespace wr22::regex_parser::parser {
11
26 regex::Part parse_regex(const utils::UnicodeStringView& regex);
27
28 }  // namespace wr22::regex_parser::parser
```

7.5 include/wr22/regex_parser/regex/capture.hpp File Reference

```
#include <wr22/regex_parser/regex/named_capture_flavor.hpp>
#include <wr22/regex_parser/utils/adt.hpp>
#include <iosfwd>
#include <string>
```

Classes

- struct wr22::regex_parser::regex::capture::None
- struct wr22::regex_parser::regex::capture::Index
- struct wr22::regex_parser::regex::capture::Name
- class wr22::regex_parser::regex::Capture

  Group capture behavior.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex
- namespace wr22::regex_parser::regex::capture

Typedefs

- using wr22::regex_parser::regex::capture::Adt = utils::Adt< None, Index, Name >

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, const Capture &capture)

7.6 capture.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // wr22
4 #include <wr22/regex_parser/regex/named_capture_flavor.hpp>
5 #include <wr22/regex_parser/utils/adt.hpp>
6
7 // stl
8 #include <iosfwd>
9 #include <string>
10
11 namespace wr22::regex_parser::regex {
12
13 class Capture;
14
15 namespace capture {
16     struct None {
17         bool operator==(const None& rhs) const = default;
18     };
19
20     struct Index {
21         bool operator==(const Index& rhs) const = default;
22     };
23
24     struct Name {
25         std::string name;
26         NamedCaptureFlavor flavor;
27         bool operator==(const Name& rhs) const = default;
28     };
29
30     using Adt = utils::Adt<None, Index, Name>;
31 }  // namespace capture
32
38 //
42 class Capture : public capture::Adt {
43 public:
44     using capture::Adt::Adt;
45 };
46
47 std::ostream& operator«(std::ostream& out, const Capture& capture);
48
49 }  // namespace wr22::regex_parser::regex
```

7.7 include/wr22/regex_parser/regex/named_capture_flavor.hpp File Reference

```
#include <iosfwd>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex

Enumerations

- enum class wr22::regex_parser::regex::NamedCaptureFlavor { wr22::regex_parser::regex::Apostrophes , wr22::regex_parser::regex::Angles , wr22::regex_parser::regex::AnglesWithP }

  The flavor (dialect) of a named group capture.

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, NamedCaptureFlavor flavor)

7.8 named_capture_flavor.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // stl
4 #include <iosfwd>
5
6 namespace wr22::regex_parser::regex {
7
12 enum class NamedCaptureFlavor
13 {
16     Apostrophes,
19     Angles,
21     AnglesWithP,
22 };
23
24 std::ostream& operator«(std::ostream& out, NamedCaptureFlavor flavor);
25
26 }  // namespace wr22::regex_parser::regex
```

7.9 include/wr22/regex_parser/regex/part.hpp File Reference

```
#include <wr22/regex_parser/regex/capture.hpp>
#include <wr22/regex_parser/utils/adt.hpp>
#include <iosfwd>
#include <memory>
#include <vector>
```

Classes

- struct wr22::regex_parser::regex::part::Empty

  An empty regex part.
- struct wr22::regex_parser::regex::part::Literal

  An regex part that matches a single character literally.
- struct wr22::regex_parser::regex::part::Alternatives

  A regex part with the list of alternatives to be matched.
- struct wr22::regex_parser::regex::part::Sequence

  A regex part with the list of items to be matched one after another.
- struct wr22::regex_parser::regex::part::Group

  A regex part that represents a group in parentheses.
- class wr22::regex_parser::regex::Part

  A part of a regular expression and its AST node type.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex
- namespace wr22::regex_parser::regex::part

  The namespace with the variants of `Part`.

Typedefs

- using wr22::regex_parser::regex::part::Adt = utils::Adt< Empty, Literal, Alternatives, Sequence, Group >

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, const Part &part)

  Convert a `Part` to a textual representation and write it to an `std::ostream`.

7.10 part.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // wr22
4 #include <wr22/regex_parser/regex/capture.hpp>
5 #include <wr22/regex_parser/utils/adt.hpp>
6
7 // stl
8 #include <iosfwd>
9 #include <memory>
10 #include <vector>
11
12 namespace wr22::regex_parser::regex {
13
14 // Forward declaration of 'Part'.
15 class Part;
16
20 namespace part {
25     struct Empty {
26         bool operator==(const Empty& rhs) const = default;
27     };
28
33     struct Literal {
34         char32_t character;
35         bool operator==(const Literal& rhs) const = default;
36     };
37
46     struct Alternatives {
48         std::vector<Part> alternatives;
49         bool operator==(const Alternatives& rhs) const = default;
50     };
51
57     struct Sequence {
59         std::vector<Part> items;
60         bool operator==(const Sequence& rhs) const = default;
61     };
62
73     struct Group {
75         Group(Capture capture, Part inner);
76
78         Capture capture;
80         std::unique_ptr<Part> inner;
81         bool operator==(const Group& rhs) const;
82     };
83
84     using Adt = utils::Adt<Empty, Literal, Alternatives, Sequence, Group>;
85 }  // namespace part
86
110 class Part : public part::Adt {
111 public:
112     using part::Adt::Adt;
113 };
114
116 std::ostream& operator<<(std::ostream& out, const Part& part);
117
118 }  // namespace wr22::regex_parser::regex
```

7.11 include/wr22/regex_parser/utils/adt.hpp File Reference

```
#include <utility>
#include <variant>
```

Classes

- struct wr22::regex_parser::utils::detail::adt::MultiCallable< Fs >
- class wr22::regex_parser::utils::Adt< Variants >

  A helper class that simplifies creation of algebraic data types.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::utils
- namespace wr22::regex_parser::utils::detail
- namespace wr22::regex_parser::utils::detail::adt

Functions

- template<typename... Variants>
  bool wr22::regex_parser::utils::operator== (const Adt< Variants... > &lhs, const Adt< Variants... > &rhs)

  Compare two compatible ADTs for equality.
- template<typename... Variants>
  bool wr22::regex_parser::utils::operator!= (const Adt< Variants... > &lhs, const Adt< Variants... > &rhs)

  Compare two compatible ADTs for non-equality.

7.12 adt.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // stl
4 #include <utility>
5 #include <variant>
6
7 namespace wr22::regex_parser::utils {
8
9 namespace detail::adt {
10     // https://en.cppreference.com/w/cpp/utility/variant/visit#Example provides a very similar
11     // example of C++ template black magic.
12     template <typename... Fs>
13     struct MultiCallable : public Fs... {
14         MultiCallable(Fs&&... fs) : Fs(fs)... {}
15     };
16 }  // namespace detail::adt
17
27 template <typename... Variants>
28 class Adt {
29 public:
31     using VariantType = std::variant<Variants...>;
32
42     template <typename V>
43     Adt(V variant) : m_variant(std::move(variant)) {}
44
73     template <typename... Fs>
74     decltype(auto) visit(Fs&&... visitors) const {
75         return std::visit(
76             detail::adt::MultiCallable<Fs...>(std::forward<Fs>(visitors)...),
77             m_variant);
78     }
79
86     template <typename... Fs>
87     decltype(auto) visit(Fs&&... visitors) {
88         return std::visit(
89             detail::adt::MultiCallable<Fs...>(std::forward<Fs>(visitors)...),
90             m_variant);
91     }
92
```

```
94      const VariantType& as_variant() const {
95          return m_variant;
96      }
97
99      VariantType& as_variant() {
100         return m_variant;
101     }
102
103 protected:
104     VariantType m_variant;
105 };
106
108 template <typename... Variants>
109 bool operator==(const Adt<Variants...>& lhs, const Adt<Variants...>& rhs) {
110     return lhs.as_variant() == rhs.as_variant();
111 }
112
114 template <typename... Variants>
115 bool operator!=(const Adt<Variants...>& lhs, const Adt<Variants...>& rhs) {
116     return !(lhs == rhs);
117 }
118
119 }  // namespace wr22::regex_parser::utils
```

7.13 include/wr22/regex_parser/utils/utf8_string_view.hpp File Reference

```
#include <compare>
#include <cstddef>
#include <exception>
#include <iterator>
#include <optional>
#include <string_view>
```

Classes

- class wr22::regex_parser::utils::UnicodeStringView

  A wrapper around `std::string_view` that enables UTF-8 codepoint iteration.
- struct wr22::regex_parser::utils::UnicodeStringView::InvalidUtf8

  An error thrown when the string's encoding is not valid UTF-8.
- class wr22::regex_parser::utils::UnicodeStringViewIterator

  Iterator over code points for `UnicodeStringView`.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::utils

7.14 utf8_string_view.hpp

Go to the documentation of this file.

```
1 #pragma once
2
3 // STL
4 #include <compare>
5 #include <cstddef>
6 #include <exception>
7 #include <iterator>
8 #include <optional>
9 #include <string_view>
10
11 namespace wr22::regex_parser::utils {
12
13 class UnicodeStringViewIterator;
14
22 class UnicodeStringView {
23     friend class UnicodeStringViewIterator;
24
25 public:
27     struct InvalidUtf8 : std::exception {
28         const char* what() const noexcept override;
29     };
30
42     explicit UnicodeStringView(const std::string_view& raw);
43
45     const std::string_view& raw() const;
46
51     UnicodeStringViewIterator begin() const;
52
57     UnicodeStringViewIterator end() const;
58
59 private:
60     std::string_view m_raw;
61 };
62
68 class UnicodeStringViewIterator {
69     friend class UnicodeStringView;
70
71 public:
73     using value_type = char32_t;
74
76     using category_type = std::forward_iterator_tag;
77
79     UnicodeStringViewIterator() = delete;
80
85     UnicodeStringViewIterator& operator++();
86
91     char32_t operator*();
92
93     bool operator==(const UnicodeStringViewIterator& rhs) const;
94     bool operator!=(const UnicodeStringViewIterator& rhs) const;
95
96 private:
97     using UnderlyingIterator = std::string_view::const_iterator;
98
99     explicit UnicodeStringViewIterator(
100         const UnderlyingIterator& iter,
101         const UnderlyingIterator& end);
102     std::optional<char32_t> decode_current_codepoint();
103     void update_current_codepoint_size();
104
105     UnderlyingIterator m_iter;
106     UnderlyingIterator m_end;
107
108     // A memoization of the result of decoding of the last code point.
109     // This is done purely for optimization to avoid having to decode a code point
110     // twice: for dereferencing the iterator and for incrementing it. The value
111     // of 0 indicates an absence of a memoized value (0 is not a valid byte size of
112     // a codepoint).
113     size_t m_last_codepoint_size = 0;
114 };
115
116 }  // namespace wr22::regex_parser::utils
```

7.15 src/parser/capture.cpp File Reference

```
#include <wr22/regex_parser/regex/capture.hpp>
#include <wr22/regex_parser/regex/named_capture_flavor.hpp>
```

```
#include <iterator>
#include <ostream>
#include <boost/locale/utf.hpp>
#include <fmt/core.h>
#include <fmt/ostream.h>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, const Capture &capture)

7.16 src/parser/errors.cpp File Reference

```
#include <wr22/regex_parser/parser/errors.hpp>
#include <fmt/core.h>
#include <boost/locale/encoding_utf.hpp>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::parser
- namespace wr22::regex_parser::parser::errors

7.17 src/parser/regex.cpp File Reference

```
#include <boost/locale/utf.hpp>
#include <wr22/regex_parser/parser/errors.hpp>
#include <wr22/regex_parser/parser/regex.hpp>
#include <wr22/regex_parser/regex/part.hpp>
#include <optional>
#include <stdexcept>
#include <string>
#include <vector>
#include <boost/locale/encoding_utf.hpp>
```

Classes

- class wr22::regex_parser::parser::Parser< Iter, Sentinel >

  A regex parser.

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::parser

Functions

- template<typename Iter , typename Sentinel >
  wr22::regex_parser::parser::Parser (Iter begin, Sentinel end) -> Parser< Iter, Sentinel >

  The type deduction guideline for `Parser`.
- regex::Part wr22::regex_parser::parser::parse_regex (const utils::UnicodeStringView &regex)

  Parse a regular expression into its AST.

7.18 src/regex/named_capture_flavor.cpp File Reference

```
#include <wr22/regex_parser/regex/named_capture_flavor.hpp>
#include <ostream>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, NamedCaptureFlavor flavor)

7.19 src/regex/part.cpp File Reference

```
#include <wr22/regex_parser/regex/part.hpp>
#include <iterator>
#include <ostream>
#include <boost/locale/utf.hpp>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::regex

Functions

- std::ostream & wr22::regex_parser::regex::operator<< (std::ostream &out, const Part &part)

  Convert a `Part` to a textual representation and write it to an `std::ostream`.

7.20 src/utils/utf8_string_view.cpp File Reference

```
#include <wr22/regex_parser/utils/utf8_string_view.hpp>
#include <boost/locale/utf.hpp>
#include <cassert>
```

Namespaces

- namespace wr22
- namespace wr22::regex_parser
- namespace wr22::regex_parser::utils

Typedefs

- using wr22::regex_parser::utils::utf_traits = boost::locale::utf::utf_traits< char >

Index