

The syntax dialect: creating a parser generator with MLIR

Fabian Mora-Cordero (U. of Delaware, USA)
fmorac@udel.edu

Dr. Sunita Chandrasekaran (U. of Delaware, USA)

Thursday 24th October, 2024

Outline

- 1 Motivation
- 2 The syntax dialect
- 3 Lexical analysis with syntax
- 4 Syntax analysis with syntax
- 5 Future work: JIT compiling syntax
- 6 Summary

Motivation

- My PhD dissertation explored how to create more modular compilers and languages
- Leading the work to explore dynamic parsing combinators for easily extending language syntax
- None of the existing tools provided me the required flexibility
- So I built one

```

1 fn [extern] omp_get_thread_num() -> i32;
2 fn main() {
3   let bsz: i32 = 4; // Block size
4   let gsz: i32 = 2; // Grid size
5   omp parallel firstprivate(bsz, gsz) {
6     let ompId = omp_get_thread_num();
7     gpu::region<<[bsz], [gsz]>> {
8       let tid : i32 = threadIdx.x;
9       let bid : i32 = blockIdx.x;
10      mlir::inline(ompId: 'i32', tid: 'i32', bid: 'i32') '
11        ',
12      gpu.printf "Host Thread ID: %d, Block ID: %d, Thread
13        ID: %d\n" %ompId, %bid, %tid : i32, i32, i32
14      ''';
15    }
16  }
17 }

```

Program for printing thread ID information from a GPU

Motivation

- My PhD dissertation explored how to create more modular compilers and languages
- Leading the work to explore dynamic parsing combinators for easily extending language syntax
- None of the existing tools provided me the required flexibility
- So I built one

```

1 fn [extern] omp_get_thread_num() -> i32;
2 fn main() {
3   let bsz: i32 = 4; // Block size
4   let gsz: i32 = 2; // Grid size
5   omp parallel firstprivate(bsz, gsz) {
6     let ompId = omp_get_thread_num();
7     gpu::region<<[bsz], [gsz]>> {
8       let tid : i32 = threadIdx.x;
9       let bid : i32 = blockIdx.x;
10      mlir::inline(ompId: 'i32', tid: 'i32', bid: 'i32') '
11        ',
12      gpu.printf "Host Thread ID: %d, Block ID: %d, Thread
13        ID: %d\n" %ompId, %bid, %tid : i32, i32, i32
14      ''';
15    }
16  }
17 }

```

Program for printing thread ID information from a GPU

Motivation

- My PhD dissertation explored how to create more modular compilers and languages
- Leading the work to explore dynamic parsing combinators for easily extending language syntax
- **None of the existing tools provided me the required flexibility**
- So I built one

```

1 fn [extern] omp_get_thread_num() -> i32;
2 fn main() {
3   let bsz: i32 = 4; // Block size
4   let gsz: i32 = 2; // Grid size
5   omp parallel firstprivate(bsz, gsz) {
6     let ompId = omp_get_thread_num();
7     gpu::region<<[bsz], [gsz]>> {
8       let tid : i32 = threadIdx.x;
9       let bid : i32 = blockIdx.x;
10      mlir::inline(ompId: 'i32', tid: 'i32', bid: 'i32') '
11        ',
12      gpu.printf "Host Thread ID: %d, Block ID: %d, Thread
13                ID: %d\n" %ompId, %bid, %tid : i32, i32, i32
14      ''';
15    }
16  }
17 }

```

Program for printing thread ID information from a GPU

Motivation

- My PhD dissertation explored how to create more modular compilers and languages
- Leading the work to explore dynamic parsing combinators for easily extending language syntax
- None of the existing tools provided me the required flexibility
- **So I built one**

```

1 fn [extern] omp_get_thread_num() -> i32;
2 fn main() {
3   let bsz: i32 = 4; // Block size
4   let gsz: i32 = 2; // Grid size
5   omp parallel firstprivate(bsz, gsz) {
6     let ompId = omp_get_thread_num();
7     gpu::region<<[bsz], [gsz]>> {
8       let tid : i32 = threadIdx.x;
9       let bid : i32 = blockIdx.x;
10      mlir::inline(ompId: 'i32', tid: 'i32', bid: 'i32') '
11        ',
12      gpu.printf "Host Thread ID: %d, Block ID: %d, Thread
13                ID: %d\n" %ompId, %bid, %tid : i32, i32, i32
14      ''';
15    }
16  }
17 }

```

Program for printing thread ID information from a GPU

The syntax dialect: IR

- Provides a high-level description of formal grammars
- Incorporates useful high-level constructs like functions over syntax expressions
- Has dedicated operations for lexical and syntax analysis

```

1  syntax.macro @Interleave(%arg0: !syntax.expr,
2     %arg1: !syntax.expr) -> !syntax.expr {
3     %and = and %arg1, %arg0
4     %zom = zero_or_more %and
5     %and_0 = and %arg0, %zom
6     return %and_0
7 }
8 // a(,a)*
9 syntax.rule @rule {
10 %a = terminal #syntax.literal<"a">
11 %comma = terminal #syntax.literal<",">
12 %ret = call @Interleave(%a, %comma)
13 return %ret
14 }
15 // mlir-opt --inline
16 syntax.rule @rule{
17 %a = terminal #syntax.literal<"a">
18 %comma = terminal #syntax.literal<",">
19 %and = and %a, %comma
20 %zom = zero_or_more %and
21 %and_1 = and %a, %zom
22 return %and_1
23 }

```

Interleave macro example

The syntax dialect: IR

- Provides a high-level description of formal grammars
- Incorporates useful high-level constructs like functions over syntax expressions
- Has dedicated operations for lexical and syntax analysis

```

1  syntax.macro @Interleave(%arg0: !syntax.expr,
2     %arg1: !syntax.expr) -> !syntax.expr {
3     %and = and %arg1, %arg0
4     %zom = zero_or_more %and
5     %and_0 = and %arg0, %zom
6     return %and_0
7 }
8 // a(,a)*
9 syntax.rule @rule {
10 %a = terminal #syntax.literal<"a">
11 %comma = terminal #syntax.literal<",">
12 %ret = call @Interleave(%a, %comma)
13 return %ret
14 }
15 // mlir-opt --inline
16 syntax.rule @rule{
17 %a = terminal #syntax.literal<"a">
18 %comma = terminal #syntax.literal<",">
19 %and = and %a, %comma
20 %zom = zero_or_more %and
21 %and_1 = and %a, %zom
22 return %and_1
23 }

```

Interleave macro example

The syntax dialect: IR

- Provides a high-level description of formal grammars
- Incorporates useful high-level constructs like functions over syntax expressions
- Has dedicated operations for lexical and syntax analysis

```

1  syntax.macro @Interleave(%arg0: !syntax.expr,
2     %arg1: !syntax.expr) -> !syntax.expr {
3     %and = and %arg1, %arg0
4     %zom = zero_or_more %and
5     %and_0 = and %arg0, %zom
6     return %and_0
7 }
8 // a(,a)*
9 syntax.rule @rule {
10 %a = terminal #syntax.literal<"a">
11 %comma = terminal #syntax.literal<",">
12 %ret = call @Interleave(%a, %comma)
13 return %ret
14 }
15 // mlir-opt --inline
16 syntax.rule @rule{
17 %a = terminal #syntax.literal<"a">
18 %comma = terminal #syntax.literal<",">
19 %and = and %a, %comma
20 %zom = zero_or_more %and
21 %and_1 = and %a, %zom
22 return %and_1
23 }

```

Interleave macro example

The syntax dialect: IR

- Most operations are Pure or constant-like
- We can use CSE to optimize redundancy

```

1 // Rule for (ab | abc)
2 syntax.rule @rule {
3   %a = terminal #syntax.literal<"a">
4   %b = terminal #syntax.literal<"b">
5   %c = terminal #syntax.literal<"c">
6   %ab = and %a, %b
7   %tmp = and %a, %b
8   %abc = and %tmp, %c
9   %ret = or %ab, %abc
10  return %ret
11 }
12 // mlir-opt --cse
13 syntax.rule @rule{
14   %a = terminal #syntax.literal<"a">
15   %b = terminal #syntax.literal<"b">
16   %c = terminal #syntax.literal<"c">
17   %and = and %a, %b
18   %and_2 = and %and, %c
19   %or = or %and, %and_2
20   return %or
21 }

```

Employing CSE for grammar simplification

The syntax dialect: IR

- Most operations are Pure or constant-like
- We can use CSE to optimize redundancy

```

1 // Rule for (ab | abc)
2 syntax.rule @rule {
3   %a = terminal #syntax.literal<"a">
4   %b = terminal #syntax.literal<"b">
5   %c = terminal #syntax.literal<"c">
6   %ab = and %a, %b
7   %tmp = and %a, %b
8   %abc = and %tmp, %c
9   %ret = or %ab, %abc
10  return %ret
11 }
12 // mlir-opt --cse
13 syntax.rule @rule{
14   %a = terminal #syntax.literal<"a">
15   %b = terminal #syntax.literal<"b">
16   %c = terminal #syntax.literal<"c">
17   %and = and %a, %b
18   %and_2 = and %and, %c
19   %or = or %and, %and_2
20   return %or
21 }

```

Employing CSE for grammar simplification

Lexical analysis with syntax

Generating lexers: TableGen backend

- We created a TableGen backend to construct lexical analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The lexer generator is based on traditional DFA's

```

1 // Rule in TableGen
2 def IdentifierStx :
3   Rule<"Identifier", "[a-zA-Z_] [a-zA-Z_0-9]*">;
4 // MLIR generated by the tablegen backend
5 syntax.dfa @Lexer {
6   rule @Identifier {
7     %t0 = terminal #syntax.char_class<"A-Z_a-z">
8     %t1 = terminal #syntax.char_class<"0-9A-Z_a-z">
9     %zom = zero_or_more %t1
10    %a = and %t0, %zom
11    return %a
12  }
13 }

```

Lexer for an identifier

Generating lexers: TableGen backend

- We created a TableGen backend to construct lexical analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The lexer generator is based on traditional DFA's

```

1 // Rule in TableGen
2 def IdentifierStx :
3   Rule<"Identifier", "[a-zA-Z_] [a-zA-Z_0-9]*">;
4 // MLIR generated by the tablegen backend
5 syntax.dfa @Lexer {
6   rule @Identifier {
7     %t0 = terminal #syntax.char_class<"A-Z_a-z">
8     %t1 = terminal #syntax.char_class<"0-9A-Z_a-z">
9     %zom = zero_or_more %t1
10    %a = and %t0, %zom
11    return %a
12  }
13 }

```

Lexer for an identifier

Generating lexers: TableGen backend

- We created a TableGen backend to construct lexical analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The lexer generator is based on traditional DFA's

```

1 // Rule in TableGen
2 def IdentifierStx :
3   Rule<"Identifier", "[a-zA-Z_] [a-zA-Z_0-9]*">;
4 // MLIR generated by the tablegen backend
5 syntax.dfa @Lexer {
6   rule @Identifier {
7     %t0 = terminal #syntax.char_class<"A-Z_a-z">
8     %t1 = terminal #syntax.char_class<"0-9A-Z_a-z">
9     %zom = zero_or_more %t1
10    %a = and %t0, %zom
11    return %a
12  }
13 }

```

Lexer for an identifier

Lexical transformations

- We begin with inlining, canonicalizing, and applying CSE
- We transform the IR into a DFA
- We perform traditional DFA minimization
- We generate C++ for the IR

```

1 syntax.dfa @Main {
2   rule @Identifier {
3     %t0 = terminal #syntax.char_class<"A-Z_a-z">
4     %t1 = terminal #syntax.char_class<"0-9A-Z_a-z">
5     %zom = zero_or_more %t1
6     %a = and %t0, %zom
7     return %a
8   }
9 }

```

Lexer for an identifier

Lexical transformations

- We begin with inlining, canonicalizing, and applying CSE
- **We transform the IR into a DFA**
- We perform traditional DFA minimization
- We generate C++ for the IR

```

1 dfa @Main {
2   %eps = eps
3   %terminal = terminal #syntax.char_class<"A-Z_a-z">
4   %terminal_0 = terminal #syntax.char_class<"0-9A-Z_a-z">
5   lex_state @StateMain0 {
6     lex_transition %terminal -> @StateMain1
7   }
8   lex_state @StateMain1 {
9     lex_transition %terminal_0 -> @StateMain1
10    lex_transition %eps -> @StateMain2
11  }
12  lex_state final @StateMain2 id = @Identifier
13 }

```

Code after minimizing the DFA

Lexical transformations

- We begin with inlining, canonicalizing, and applying CSE
- We transform the IR into a DFA
- **We perform traditional DFA minimization**
- We generate C++ for the IR

```

1 dfa @Main {
2   %eps = eps
3   %terminal = terminal #syntax.char_class<"A-Z_a-z">
4   %terminal_0 = terminal #syntax.char_class<"0-9A-Z_a-z">
5   lex_state @StateMain0 {
6     lex_transition %terminal -> @StateMain1
7   }
8   lex_state @StateMain1 {
9     lex_transition %terminal_0 -> @StateMain1
10    lex_transition %eps -> @StateMain2
11  }
12  lex_state final @StateMain2 id = @Identifier
13 }

```

Code after minimizing the DFA

Lexical transformations

- We begin with inlining, canonicalizing, and applying CSE
- We transform the IR into a DFA
- We perform traditional DFA minimization
- We generate C++ for the IR

```

1  Lexer::TokenID
2  Lexer::lexMain(SourceState &state, SourceLocation &beginLoc,
3                llvm::StringRef &spelling) const {
4  StateMain0: {
5      if ((/*A*/ (65 <= *state) && (*state <= 90) /*Z*/) ||
6          *state == 95 /*_*/ ||
7          (/*a*/ (97 <= *state) && (*state <= 122) /*z*/)) {
8          state.advance();
9          goto StateMain1;
10     }
11     return Invalid;
12 }
13 StateMain1: {
14     if ((/*0*/ (48 <= *state) && (*state <= 57) /*9*/) ||
15         (/*A*/ (65 <= *state) && (*state <= 90) /*Z*/) ||
16         *state == 95 /*_*/ ||
17         (/*a*/ (97 <= *state) && (*state <= 122) /*z*/)) {
18         state.advance();
19         goto StateMain1;
20     }
21     spelling = getSpelling(beginLoc, state.getLoc());
22     return Identifier;
23 }
24     return Invalid;
25 }

```

Syntax analysis with syntax

Generating parsers: TableGen backend

- We created a TableGen backend to construct syntax analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The TableGen backend generates a Packrat-PEG parser
- It has native support for dynamic combinators

```

1 def Dim3: ParserMacro<"Dim3", ["expr"], [{
2   expr ("," expr ("," expr)? )?
3 }]>;
4 /*example:
5 region<<[n], [32, 16]>> {
6   // statement
7 }
8 */
9 def GPU_RegionStx:
10  Production<"Region", "::mlir::Operation*"> {
11  let rule = [{
12    /*parse launch bounds*/
13    "region" "<<"
14    "[" @Dim3(#dyn("Expr")):$bv { bsz.push_back(
15      bv.get()); }]"
16    ","
17    "[" @Dim3(#dyn("Expr")):$gv { gsz.push_back(
18      gv.get()); }]"
19    ">>"
20    /*parse region statement*/
21    #dyn("Stmt")
22  }];

```

TableGen code for parsing a GPU region

Generating parsers: TableGen backend

- We created a TableGen backend to construct syntax analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The TableGen backend generates a Packrat-PEG parser
- It has native support for dynamic combinators

```

1 def Dim3: ParserMacro<"Dim3", ["expr"], [{
2   expr ("," expr ("," expr)? )?
3 }]>;
4 /*example:
5 region<<[n], [32, 16]>> {
6   // statement
7 }
8 */
9 def GPU_RegionStx:
10  Production<"Region", "::mlir::Operation*"> {
11  let rule = [{
12    /*parse launch bounds*/
13    "region" "<<"
14    "[" @Dim3(#dyn("Expr")):$bv { bsz.push_back(
15      bv.get()); }]"
16    ","
17    "[" @Dim3(#dyn("Expr")):$gv { gsz.push_back(
18      gv.get()); }]"
19    ">>"
20    /*parse region statement*/
21    #dyn("Stmt")
22  }];

```

TableGen code for parsing a GPU region

Generating parsers: TableGen backend

- We created a TableGen backend to construct syntax analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- **The TableGen backend generates a Packrat-PEG parser**
- It has native support for dynamic combinators

```

1  def Dim3: ParserMacro<"Dim3", ["expr"], [{
2      expr ("," expr ("," expr)? )?
3  }]>;
4  /*example:
5  region<<[n], [32, 16]>> {
6      // statement
7  }
8  */
9  def GPU_RegionStx:
10     Production<"Region", "::mlir::Operation*"> {
11     let rule = [{
12         /*parse launch bounds*/
13         "region" "<<"
14         "[" @Dim3(#dyn("Expr")):$bv { bsz.push_back(
15             bv.get()); }> "]"
16         ","
17         "[" @Dim3(#dyn("Expr")):$gv { gsz.push_back(
18             gv.get()); }> "]"
19         ">>"
20         /*parse region statement*/
21         #dyn("Stmt")
22     }];
23 }

```

TableGen code for parsing a GPU region

Generating parsers: TableGen backend

- We created a TableGen backend to construct syntax analyzers
- The MLIR code then gets analyzed and optimized and then translated to C++
- The TableGen backend generates a Packrat-PEG parser
- It has native support for dynamic combinators

```

1 def Dim3: ParserMacro<"Dim3", ["expr"], [{
2   expr ("," expr ("," expr)? )?
3 }]>;
4 /*example:
5 region<<[n], [32, 16]>> {
6   // statement
7 }
8 */
9 def GPU_RegionStx:
10  Production<"Region", "::mlir::Operation*"> {
11  let rule = {
12    /*parse launch bounds*/
13    "region" "<<"
14    "[" @Dim3(#dyn("Expr")):$bv { bsz.push_back(
15      bv.get()); }> "]"
16    ","
17    "[" @Dim3(#dyn("Expr")):$gv { gsz.push_back(
18      gv.get()); }> "]"
19    ">>"
20    /*parse region statement*/
21    #dyn("Stmt")
22  };
23 }

```

TableGen code for parsing a GPU region

Syntax transformations

- We take our TableGen spec and transform it into syntax
- We then perform inlining, canonicalize, and apply CSE
- We analyze the grammar and lower its complexity
- We generate C++ for the IR

```

1 def Top : Production<"Top", "::mlir::Value"> {
2   let rule = [{
3     ("a" "b") |
4     ("a" "c") |
5     ("c" "d")
6   }];
7 }
8 def Syntax : Parser<"Syntax", XBLangLexer> {
9   let macros = [];
10  let productions = [
11    Top,
12  ];
13  let startSymbol = "Top";
14  let cppNamespace = "::";
15  let defaultToken = "Identifier";
16 }

```

Syntax specification

Syntax transformations

- We take our TableGen spec and transform it into `syntax`
- We then perform inlining, canonicalize, and apply CSE
- We analyze the grammar and lower its complexity
- We generate C++ for the IR

```

1  syntax.parser @Syntax start = @Top {
2  rule @Top {
3  %Identifier = terminal #syntax.lex_terminal<@Identifier, unk, "a">
4  %md = md_node %Identifier @_0
5  %Identifier_0 = terminal #syntax.lex_terminal<@Identifier, unk, "b">
6  %md_1 = md_node %Identifier_0 @_1
7  %and = and %md, %md_1
8  %Identifier_2 = terminal #syntax.lex_terminal<@Identifier, unk, "a">
9  %md_3 = md_node %Identifier_2 @_0
10 %Identifier_4 = terminal #syntax.lex_terminal<@Identifier, unk, "c">
11 %md_5 = md_node %Identifier_4 @_1
12 %and_6 = and %md_3, %md_5
13 %or = or %and, %and_6
14 %Identifier_7 = terminal #syntax.lex_terminal<@Identifier, unk, "c">
15 %md_8 = md_node %Identifier_7 @_0
16 %Identifier_9 = terminal #syntax.lex_terminal<@Identifier, unk, "d">
17 %md_10 = md_node %Identifier_9 @_1
18 %and_11 = and %md_8, %md_10
19 %or_12 = or %or, %and_11
20 return %or_12
21 }
22 }

```

Generated syntax code

Syntax transformations

- We take our TableGen spec and transform it into `syntax`
- We then perform inlining, canonicalize, and apply CSE
- We analyze the grammar and lower its complexity
- We generate C++ for the IR

```

1  syntax.parser @Syntax start = @Top {
2    rule @Top {
3      %Identifier = terminal #syntax.lex_terminal<@Identifier, unk, "a">
4      %md = md_node %Identifier @_0
5      %Identifier_0 = terminal #syntax.lex_terminal<@Identifier, unk, "b">
6      %md_1 = md_node %Identifier_0 @_1
7      %and = and %md, %md_1
8      %Identifier_2 = terminal #syntax.lex_terminal<@Identifier, unk, "a">
9      %md_3 = md_node %Identifier_2 @_0
10     %Identifier_4 = terminal #syntax.lex_terminal<@Identifier, unk, "c">
11     %md_5 = md_node %Identifier_4 @_1
12     %and_6 = and %md_3, %md_5
13     %or = or %and, %and_6
14     %Identifier_7 = terminal #syntax.lex_terminal<@Identifier, unk, "c">
15     %md_8 = md_node %Identifier_7 @_0
16     %Identifier_9 = terminal #syntax.lex_terminal<@Identifier, unk, "d">
17     %md_10 = md_node %Identifier_9 @_1
18     %and_11 = and %md_8, %md_10
19     %or_12 = or %or, %and_11
20     return %or_12
21   }
22 }

```

Generated syntax code

Syntax transformations

- We take our TableGen spec and transform it into `syntax`
- We then perform inlining, canonicalize, and apply CSE
- We analyze the grammar and lower its complexity
- We generate C++ for the IR

```

1  rule @Top attributes {first_set = [#syntax.lex_terminal<@Identifier,
   unk, "c">, #syntax.lex_terminal<@Identifier, unk, "a">]} {
2  %Identifier = terminal #syntax.lex_terminal<@Identifier, unk, "d">
3  %Identifier_0 = terminal #syntax.lex_terminal<@Identifier, unk, "c">
4  %Identifier_1 = terminal #syntax.lex_terminal<@Identifier, unk, "b">
5  %Identifier_2 = terminal #syntax.lex_terminal<@Identifier, unk, "a">
6  %md = md_node %Identifier_2 @_0
7  %md_3 = md_node %Identifier_1 @_1
8  %seq = seq %md, %md_3
9  %md_4 = md_node %Identifier_2 @_0
10 %md_5 = md_node %Identifier_0 @_1
11 %seq_6 = seq %md_4, %md_5
12 %md_7 = md_node %Identifier_0 @_0
13 %md_8 = md_node %Identifier @_1
14 %seq_9 = seq %md_7, %md_8
15 %any = any %seq, %seq_6 first_sets = [[#syntax.lex_terminal<
   @Identifier, unk, "a">], [#syntax.lex_terminal<@Identifier, unk,
   "a">]] conflicts = [[#syntax.lex_terminal<@Identifier, unk, "
   a">], [#syntax.lex_terminal<@Identifier, unk, "a">]]
16 %switch = switch %any, %seq_9 first_sets = [[#syntax.lex_terminal<
   @Identifier, unk, "a">], [#syntax.lex_terminal<@Identifier, unk,
   "c">]]
17 return %switch
18 }

```

Future work: JIT compiling syntax

- We are currently working on JIT compiling **syntax**
 - Allowing the creation of LLVM-optimized regex recognizers on the fly
 - Allowing the creation of self-extensible languages
- The biggest barrier for JIT compilation is transforming the associated runtime library to be less C++ reliant

Future work: JIT compiling syntax

- We are currently working on JIT compiling **syntax**
 - **Allowing the creation of LLVM-optimized regex recognizers on the fly**
 - Allowing the creation of self-extensible languages
- The biggest barrier for JIT compilation is transforming the associated runtime library to be less C++ reliant

Future work: JIT compiling syntax

- We are currently working on JIT compiling **syntax**
 - Allowing the creation of LLVM-optimized regex recognizers on the fly
 - **Allowing the creation of self-extensible languages**
- The biggest barrier for JIT compilation is transforming the associated runtime library to be less C++ reliant

Future work: JIT compiling syntax

- We are currently working on JIT compiling **syntax**
 - Allowing the creation of LLVM-optimized regex recognizers on the fly
 - Allowing the creation of self-extensible languages
- The biggest barrier for JIT compilation is transforming the associated runtime library to be less C++ reliant

Summary

- **syntax** allows to represent and create lexical analyzers
- `syntax` can represent syntax analyzers
- It shows how traditional techniques such as CSE can be applied in non-traditional contexts
- This approach proves once more how MLIR can be used in many areas beyond traditional compilation

Summary

- `syntax` allows to represent and create lexical analyzers
- `syntax` can represent syntax analyzers
- It shows how traditional techniques such as CSE can be applied in non-traditional contexts
- This approach proves once more how MLIR can be used in many areas beyond traditional compilation

Summary

- `syntax` allows to represent and create lexical analyzers
- `syntax` can represent syntax analyzers
- It shows how traditional techniques such as CSE can be applied in non-traditional contexts
- This approach proves once more how MLIR can be used in many areas beyond traditional compilation

Summary

- `syntax` allows to represent and create lexical analyzers
- `syntax` can represent syntax analyzers
- It shows how traditional techniques such as CSE can be applied in non-traditional contexts
- This approach proves once more how MLIR can be used in many areas beyond traditional compilation

Acknowledgments

Acknowledgments

- This material is based upon work supported by the National Science Foundation (NSF) under grant no. 1814609.
- This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.
- This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.
- This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.



Questions?