Analyzing the Clang diagnostics experience

Matt Beaumont-Gay
matthewbg@google.com

Agenda

1. A brief introduction to Clang diagnostics.
2. The design and implementation of the diagnostics processor.
3. Statistics and anecdotes.
4. Q&A.

Anatomy of a Clang diagnostic

include/clang/Basic/DiagnosticSemaKinds.td:

```
def warn_no_constructor_for_refconst : Warning<
 "%select{struct|union|class|enum}0 %1 does not declare any
constructor to initialize its non-modifiable members">;
```

Anatomy of a Clang diagnostic

include/clang/Basic/DiagnosticSemaKinds.td:

```
def warn_no_constructor_for_refconst : Warning<
 "%select{struct|union|class|enum}0 %1 does not declare any
constructor to initialize its non-modifiable members">;
```

lib/Sema/SemaDeclCXX.cpp:

```
if (F->getType()->isReferenceType() ||
    (F->getType().isConstQualified() &&
     F->getType()->isScalarType())) {
  Diag(Record->getLocation(),
       diag::warn_no_constructor_for_refconst)
      << Record->getTagKind() << Record;
```

The Goal

```
path/to/file.cc:42:5: error: class 'Foo' does not
declare any constructor to initialize its non-
modifiable members
```

The Goal

path/to/file.cc:42:5: error: class 'Foo' does not declare any constructor to initialize its non-modifiable members

```
{ file: "path/to/file.cc"
  line: 42
  column: 5
  level: ERROR
  format: "%select{class|...}0 %1 does not ..."
  arguments: "class"
  arguments: "'Foo'"
  text: "class 'Foo' does not ..." }
```

But you can just ask Clang for structured output!

Yes, but...

... not an option for javac, gold, etc.

... hard to teach our build system to fetch and store a side output.

Parsing Clang diagnostics, the interesting part

include/clang/Basic/DiagnosticSemaKinds.td:

```
def warn_no_constructor_for_refconst : Warning<
 "%select{struct|union|class|enum}0 %1 does not declare any
constructor to initialize its non-modifiable members">;
```

Parsing Clang diagnostics, the interesting part

include/clang/Basic/DiagnosticSemaKinds.td:

```
def warn_no_constructor_for_refconst : Warning<
 "%select{struct|union|class|enum}0 %1 does not declare any
constructor to initialize its non-modifiable members">;
```

```
(struct|union|class|enum) (.*) does not declare any...
```

Parsing Clang diagnostics, the interesting part

```
"unknown type name %0; did you mean %1?"
```

```
"unknown type name %0"
```

```
"%0"
```

Parsing Clang diagnostics, the interesting part

```
"unknown type name %0; did you mean %1?"
unknown type name (.*); did you mean (.*)\?
```

```
"unknown type name %0"
unknown type name (.*)
```

```
"%0"
(.*)
```

Parsing Clang diagnostics, the interesting part

x.c:1:1: error: unknown type name 'Bar'; did you mean 'Baz'?

```
"unknown type name %0; did you mean %1?"
unknown type name (.*); did you mean (.*)\?
```

```
"unknown type name %0"
unknown type name (.*)
```

```
"%0"
(.*)
```

Parsing Clang diagnostics, the interesting part

```
x.c:1:1: error: unknown type name 'Bar'; did you mean 'Baz'?
```

```
"unknown type name %0; did you mean %1?"
unknown type name (.*); did you mean (.*)\?
["'Bar'", "'Baz'"]
```

```
"unknown type name %0"
unknown type name (.*)
["'Bar'; did you mean 'Baz'?"]
```

```
"%0"
(.*)
["unknown type name 'Bar'; did you mean 'Baz'?"]
```

Parsing Clang diagnostics, the slightly less interesting parts

```
"(?:" ANSI_ESCAPE "([-_./<> a-zA-Z0-9]+):(\\d+):(\\d+): )?"
  ANSI_ESCAPE "(note|warning|error|fatal error): "
  ANSI_ESCAPE "(.*)" ANSI_ESCAPE
```

Machinery for recognizing include stacks, skipping over snippets and caret lines, associating notes with the primary diagnostic, ...

Behold the power of this fully armed and operational MapReduce

Map: (Build info, compiler stderr) ->

    N * (hash of diagnostic info, diagnostic info)

Reduce: Deduplicate, write out to Dremel

The Data

30,000 engineer-weeks
3,891,814 diagnostics

1,472,752 notes
591 warnings
2,383,858 errors (147,893 from `-Werror`)
34,613 fatal errors

What errors do we emit?

400876	use of undeclared identifier %0
225265	no member named %0 in %1
183930	use of undeclared identifier %0; did you mean %1?
120738	unknown type name %0; did you mean %1?
88148	unknown type name %0
72815	no member named %0 in %1; did you mean %2?
55083	no matching function for call to %0
52739	no matching constructor for initialization of %0
49759	no viable conversion%diff{ from $ to $|}0,1
44670	member access into incomplete type %0

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

24%

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

23%

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

31%

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

4%

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

57%

Typo correction FTW

```
     5 | field designator %0 does not refer to any field in type %1
     2 | field designator %0 does not refer to any field in type %1; did you mean %2?
225265 | no member named %0 in %1
 72815 | no member named %0 in %1; did you mean %2?
 25837 | no type named %0 in %1
  8148 | no type named %0 in %1; did you mean %2?
 34817 | out-of-line definition of %0 does not match any declaration in %1
  1733 | out-of-line definition of %0 does not match any declaration in %1; did you mean %2?
 88148 | unknown type name %0
120738 | unknown type name %0; did you mean %1?
400876 | use of undeclared identifier %0
183930 | use of undeclared identifier %0; did you mean %1?
```

34% have a nested-name-specifier correction

What warnings do we emit?

46128	-Wnon-virtual-dtor
25718	-Wc++98-compat
15440	-Wthread-safety-analysis
8826	-Wreturn-type
7432	-Wdelete-incomplete
6736	-Wformat
5167	-Wnon-pod-varargs
3355	-Wreorder
2457	-Wreturn-stack-address
2299	-Wvexing-parse

What warnings do we emit?

46128	-Wnon-virtual-dtor
25718	-Wc++98-compat
15440	-Wthread-safety-analysis
8826	-Wreturn-type
7432	-Wdelete-incomplete
6736	-Wformat
5167	-Wnon-pod-varargs
3355	-Wreorder
2457	-Wreturn-stack-address
2299	-Wvexing-parse

13932 from delesley

What warnings do we emit?

46128	-Wnon-virtual-dtor
25718	-Wc++98-compat
15440	-Wthread-safety-analysis
8826	-Wreturn-type
7432	-Wdelete-incomplete
6736	-Wformat
5167	-Wnon-pod-varargs
3355	-Wreorder
2457	-Wreturn-stack-address
2299	-Wvexing-parse
2111	-Woverloaded-virtual

What warnings do we emit?

46128	-Wnon-virtual-dtor
25718	-Wc++98-compat
15440	-Wthread-safety-analysis
8826	-Wreturn-type
7432	-Wdelete-incomplete
6736	-Wformat
5167	-Wnon-pod-varargs
3355	-Wreorder
2457	-Wreturn-stack-address
2299	-Wvexing-parse
2111	-Woverloaded-virtual

All from blaikie

What warnings do we emit?

46128	-Wnon-virtual-dtor
25718	-Wc++98-compat
15440	-Wthread-safety-analysis
8826	-Wreturn-type
7432	-Wdelete-incomplete
6736	-Wformat
5167	-Wnon-pod-varargs
3355	-Wreorder
2457	-Wreturn-stack-address
2299	-Wvexing-parse
2111	-Woverloaded-virtual
1820	-Winvalid-pp-token

printf, or, the worst libc API

`-Wformat, -Wformat-security, -Wnon-pod-varargs`

```
5874 | format specifies type %0 but the argument has type %1
4929 | cannot pass %select{non-POD|non-trivial}0 object of type %1 to variadic %select
{function|block|method|constructor}2; expected type from format string was %3
1466 | format string is not a string literal (potentially insecure)
 587 | more '%%' conversions than data arguments
 238 | cannot pass object of %select{non-POD|non-trivial}0 type %1 through variadic %
select{function|block|method|constructor}2; call will abort at runtime
 136 | field %select{width|precision}0 should have type %1, but argument has type %2
 102 | incomplete format specifier
  13 | flag '%0' results in undefined behavior with '%1' conversion specifier
  11 | data argument position '%0' exceeds the number of data arguments (%1)
  11 | length modifier '%0' results in undefined behavior or no effect with '%1'
conversion specifier
   2 | format string contains '\0' within the string body
```

printf, or, the worst libc API

```
format specifies type %0 but the argument has type %1

|   # | %0              | %1                                  |
+-----+-----------------+-------------------------------------+
| 507 | 'int'           | 'size_type' (aka 'unsigned long') |
| 361 | 'char *'        | 'StringPiece'                       |
| 324 | 'int'           | 'int64' (aka 'long long')           |
| 240 | 'int'           | 'double'                            |
| 174 | 'long long'     | 'int'                               |
| 168 | 'long'          | 'size_type' (aka 'unsigned int')  |
| 146 | 'int'           | 'unsigned long'                     |
| 144 | 'unsigned long' | 'size_type' (aka 'unsigned int')  |
| 138 | 'double'        | 'int'                               |
| 132 | 'unsigned int'  | 'size_type' (aka 'unsigned long') |
```

Notes. Oh, do we have notes.

avg	#	format
17.76	41	conversion %diff{from $ to $|between types}0,1 is ambiguous
11.00	22	ambiguous partial specializations of %0
09.42	26	static_assert failed %0
07.93	498	use of overloaded operator '%0' is ambiguous (with operand types %1 and %2)
07.47	2616	must use '%1' tag to refer to type %0%select{| in this scope}2
05.53	8203	invalid application of '%select{sizeof|...}0' to an incomplete type %1
05.50	4	incomplete type %0 is not assignable
05.00	63	use of overloaded operator '%0' is ambiguous (operand type %1)
04.89	9	call to object of type %0 is ambiguous
04.60	20	redefinition of label %0

Notes. Oh, do we have notes.

avg	#	format
17.76	41	conversion %diff{from $ to $|between types}0,1 is ambiguous
11.00	22	ambiguous partial specializations of %0
09.42	26	static_assert failed %0
07.93	498	use of overloaded operator '%0' is ambiguous (with operand types %1 and %2)
07.47	2616	must use '%1' tag to refer to type %0%select{| in this scope}2
05.53	8203	invalid application of '%select{sizeof|...}0' to an incomplete type %1
05.50	4	incomplete type %0 is not assignable
05.00	63	use of overloaded operator '%0' is ambiguous (operand type %1)
04.89	9	call to object of type %0 is ambiguous
04.60	20	redefinition of label %0

wat?

http://code.google.com/p/re2/source/browse/re2/re2.h#746

```
class RE2::Arg {
  Arg(void*);
#define MAKE_PARSER(type,name) Arg(type* p) : arg_(p), parser_(name) { }
  MAKE_PARSER(char,               parse_char);
  MAKE_PARSER(signed char,        parse_char);
  MAKE_PARSER(unsigned char,      parse_uchar);
  MAKE_PARSER(short,              parse_short);
  MAKE_PARSER(unsigned short,     parse_ushort);
  MAKE_PARSER(int,                parse_int);
  MAKE_PARSER(unsigned int,       parse_uint);
  MAKE_PARSER(long,               parse_long);
  MAKE_PARSER(unsigned long,      parse_ulong);
  MAKE_PARSER(long long,          parse_longlong);
  MAKE_PARSER(unsigned long long, parse_ulonglong);
  MAKE_PARSER(float,              parse_float);
  MAKE_PARSER(double,             parse_double);
  MAKE_PARSER(string,             parse_string);
  MAKE_PARSER(StringPiece,        parse_stringpiece);
  template <class T> Arg(T* p);
```

Then things go horribly wrong

```
void f(const RE2::Arg&);
void g() {
  f(NULL);
}
```

Then things go horribly wrong

```
x.cc:3:5: error: conversion from 'long' to 'const RE2::Arg' is ambiguous
  f(NULL)
    ^~~~
stddef.h:47:18: note: expanded from macro 'NULL'
#    define NULL __null
                 ^~~~~~
re2.h:888:18: note: candidate constructor
inline RE2::Arg::Arg(void* p) : arg_(p), parser_(parse_null) { }
                 ^
re2.h:826:3: note: candidate constructor
  MAKE_PARSER(char,               parse_char);
  ^
re2.h:822:3: note: expanded from macro 'MAKE_PARSER'
  Arg(type* p) : arg_(p), parser_(name) { } \
  ^
re2.h:827:3: note: candidate constructor
  MAKE_PARSER(unsigned char,      parse_uchar);
  ^
re2.h:822:3: note: expanded from macro 'MAKE_PARSER'
  Arg(type* p) : arg_(p), parser_(name) { } \
  ^
re2.h:828:3: note: candidate constructor
  MAKE_PARSER(short,              parse_short);
  ^
re2.h:817:3: note: expanded from macro 'MAKE_PARSER'
  Arg(type* p) : arg_(p), parser_(name) { } \
  ^
re2.h:824:3: note: candidate constructor
  MAKE_PARSER(unsigned short,     parse_ushort);
  ^
```

[... 60 more lines, 30 total notes]

Fatal ~~errors~~ error.

```
'%0' file not found
```

(Plus some noise from one guy who used -Wfatal-errors for an afternoon.)

The long tail, or, "... but points for trying."

```
1 | '##' cannot appear at end of macro expansion
1 | 'long %0' is invalid
1 | 'main' is not allowed to be declared static
1 | ISO C requires a translation unit to contain at least one declaration
1 | cannot declare a class template with no name
1 | cannot template a using declaration
1 | conversion function cannot convert to a function type
1 | conversion function converting %0 to its base class %1 will never be used
1 | destructor cannot be declared using a %select{typedef|type alias}1 %0 of the class name
1 | differing user-defined suffixes ('%0' and '%1') in string literal concatenation
1 | invalid filename for line marker directive
1 | magnitude of floating-point constant too small for type %0; minimum is %1
1 | multiple default labels in one switch
1 | mutex '%0' is locked exclusively and shared in the same scope
1 | templates must have C++ linkage
1 | use of address-of-label extension outside of a function body
1 | variably modified type %0 cannot be used as a template argument
```

Future work

- Session analysis
- Trend tracking
- Feedback to library maintainers
- Improving typo correction heuristics

fin.