

Three Cool Things About D

Andrei Alexandrescu

Research Scientist

Facebook

What's the Deal?

PL Landscape

- Systems/Productivity languages
- Imperative/Declarative/Functional languages
- n^{th} generation languages
- Compiled/Interpreted/JITed languages
-

Sahara

- Sahara-sized desert around the crossroads of
 - High efficiency
 - Systems-level access
 - Modeling power
 - Simplicity
 - High productivity
 - Code correctness (of parallel programs, too)

Sahara

- Sahara-sized desert around the crossroads of
 - High efficiency
 - Systems-level access
 - Modeling power
 - Simplicity
 - High productivity
 - Code correctness (of parallel programs, too)
- Traditionally: focus on a couple and try to keep others under control

Sahara

- Sahara-sized desert around the crossroads of
 - High efficiency
 - Systems-level access
 - Modeling power
 - Simplicity
 - High productivity
 - Code correctness (of parallel programs, too)
- Traditionally: focus on a couple and try to keep others under control
- C++: ruler of 1-2, strong on 3, goes south after that

Sahara

- Sahara-sized desert around the crossroads of
 - High efficiency
 - Systems-level access
 - Modeling power
 - Simplicity
 - High productivity
 - Code correctness (of parallel programs, too)
- Traditionally: focus on a couple and try to keep others under control
- C++: ruler of 1-2, strong on 3, goes south after that
- *D: holistic approach toward the centroid*

Enter D

- Doesn't replace C's memory model
 - Builds *structure* on top of it
 - All of C's `stdlib` available to D
 - You can use `malloc` and `free` at full speed
 - `alloca` too if you really wish
 - Garbage collection fostered but not required
- Layered approach to safety
- Easy and powerful generic and generative capabilities
- Principled yet practical approach to correctness
- Threading without races

Compilation model

- Compiled language, JITable subset
- Context-independent declarations
- All top-level declarations are conceptually entered in parallel
- Reader needs minimal context
- Each character is looked at *once*
- Each symbol is loaded *once*

- Fastest language to compile by a large margin

This Talk

- Introduction
- Hello, Correctness
- Not one more gorram handwritten search, linked list, nearest neighbor, merge... *Ever.*
- Operator overloading

- Conclusions

Hello, Correctness

Correctness Starts With...

```
import std.stdio;
void main()
{
    writeln("Hello, world!");
}
```

- This program is *correct*.

Hello, World! Do You Copy? Over.

- C's hello world is *incorrect*:

```
#include <stdio.h>
main()
{
    printf("hello, world\n");
}
```

- What does the program return in case of success?
- What does the program return in case of failure?

The Greeting that Always Befriendeth

- C++'s hello world is *also incorrect*:

```
#include <iostream>
int main()
{
    std::cout << "Hello, world!\n";
}
```

- 20 years of progress
 - Develop a positive attitude
 - Learn more quirks in day one
 - More trouble with multithreaded I/O



**EVERYTHING
WENT
BETTER
THAN
EXPECTED**

A Litmus Test

- Read tutorial examples
 - Engineered to make the language look good
- Consider various production scenarios
 - I/O may fail
 - Input may be unreasonably large
 - API calls (incl. allocating memory) may fail
 - Input may actually contain non-ASCII characters
- What does it take to make the examples ready for production?

echo in 9 correct lines

```
#!/usr/bin/env rdm  
import std.getopt, std.stdio, std.string;  
void main(string[] args) {  
    bool noNewline;  
    getopt(args, config.passThrough,  
           config.stopOnFirstNonOption,  
           "n", &noNewline);  
    write(join(args[1 .. $], " "),  
          noNewline ? "" : "\n");  
}
```

Past approaches to correctness

- Design and implementation validation: contracts
 - Ubiquitous as idiom (`assert!`)
 - Deserve better language support
- Error handling:
 - 40 years anniversary of the MACeAEC (Movement Against Checking `errno` And Error Codes)
 - Propagating errors manually is relentless friction even for seasoned professionals
- `try/catch`, `using`, `unwind-protect` will never scale
 - (beyond cutesy examples)
 - They all affect the normal code path

Approach to correctness

- Design/implementation validation:
 - Full-bore contracts
 - Built-in `unittest`
- Error handling:
 - Modern exception handling
 - Can `throw` from *anywhere* without loss of info
 - Destructors may `throw`!
 - Writing transactional code is easy and linear
 - No need to define types
 - No need to `try`
 - Proper sequencing, not flow, determines correctness

Transactional file copy in 9 correct lines

```
#!/usr/bin/env rdmd
import std.exception, std.file;
void main(string[] args) {
    enforce(args.length == 3,
        "Usage: trcopy file1 file2");
    auto tmp = args[2] ~ ".messedup";
    scope(failure) if (exists(tmp)) remove(tmp);
    copy(args[1], tmp);
    rename(tmp, args[2]);
}
```

Truth in advertising

- The `std.file.copy` routine itself:
 - Control flow: one `if` and one `for` in 47 lines (!)
 - Either succeeds or fails successfully
 - Makes 6 calls to `enforce`
 - Has 4 `scope` statements
 - No extra aides: uses straight `fcntl`, `stat`, and even `malloc`
- Replaces a function:
 - 105 lines long
 - 9 `gotos` to 5 labels
 - 11 other control flow statements

Results may vary. Combine with diet and exercise for best results.

Genericity

What is Generic Programming?

What is Generic Programming?

- Find the representation of an algorithm using the narrowest requirements possible
 - Big- $O()$ encapsulation should be a crime in 48 states
 - Low-level efficiency important too: reduces impetus for regressing to hand-written code

What is Generic Programming?

- Find the representation of an algorithm using the narrowest requirements possible
 - Big- $O()$ encapsulation should be a crime in 48 states
 - Low-level efficiency important too: reduces impetus for regressing to hand-written code
- Define types to implement said requirements

What is Generic Programming?

- Find the representation of an algorithm using the narrowest requirements possible
 - Big- $O()$ encapsulation should be a crime in 48 states
 - Low-level efficiency important too: reduces impetus for regressing to hand-written code
- Define types to implement said requirements
- Leverage the algorithm for ultimate reuse

What is Generic Programming?

- Find the representation of an algorithm using the narrowest requirements possible
 - Big- $O()$ encapsulation should be a crime in 48 states
 - Low-level efficiency important too: reduces impetus for regressing to hand-written code
- Define types to implement said requirements
- Leverage the algorithm for ultimate reuse
- ...
- Profit!

What is Generic Programming?

- Find the representation of an algorithm using the narrowest requirements possible
 - Big- $O()$ encapsulation should be a crime in 48 states
 - Low-level efficiency important too: reduces impetus for regressing to hand-written code
 - Define types to implement said requirements
 - Leverage the algorithm for ultimate reuse
 - ...
 - Profit!
-
- Arguably one of the noblest endeavors of our métier

Generic Programming

- “Write once, instantiate anywhere”
 - Implementations tailored for enhanced interfaces are welcome
- Prefers static type information and static expansion (macro style)
- Fosters strong mathematical underpinnings
- Has tenuous relationship with binary interfaces
- Starts with algorithms, not interfaces or objects
- “Premature encapsulation is the root of all evil”

D: Generic Programming Democratized

- Accessible and powerful generic programming capabilities

```
auto min(L, R)(L lhs, R rhs) {  
    return rhs < lhs ? rhs : lhs;  
}  
auto a = min(x, y);
```

- This function is as efficient as any specialized, handwritten version (true genericity)

(Compare At

```
template <class T>
T& min(T& lhs, T& rhs) {
    return rhs < lhs ? rhs : lhs;
}

template <class T>
const T& min(const T& lhs, const T& rhs) {
    return rhs < lhs ? rhs : lhs;
}
```

- Function is incomplete
- Revised implementation, rejected proposal N2199: 175 lines, 10 types, 12 specializations)

Iterators and Ranges

- Iterators: very useful primitives for many algorithms
 - Flexible but fundamentally uncheckable
 - D's standard library introduces ranges: abstracted pairs of iterators
 - Defines superset of STL on top of ranges
 - Ranges pervade D: I/O, lazy evaluation, random numbers...
-
- google terms: "On Iteration", std.algorithm

Containers

- No container hierarchy
 - Federation of independent containers
- Containers have too much “personality”
- Futile to classify in formal interfaces
- Yet: duck typing should work
- Efficient allocation with `malloc` possible without affecting safety (upcoming article: “Sealed Containers”)
 - Trickier & more interesting than it may seem

Algorithm example: edit distance

```
size_t levenshteinDistance
    (alias equals = "a==b", R1, R2)
    (R1 s, R2 t)
    if (isForwardRange!R1 && isForwardRange!R2)
    {
        ...
    }
```

- Usual assumptions: strict equality, inputs must offer random access
- Yet algorithm has modular comparison and requires only forward access for both inputs

Edit Distance (continued)

- D strings are array of immutable characters:

```
alias immutable(char) [] string;  
alias immutable(wchar) [] wstring;  
alias immutable(dchar) [] dstring;
```

- `string`, `wstring` formally offer only bidirectional range interface
- Consequence: `levenshteinDistance` works correctly for all strings with custom comparison—no copy, no handcrafting!
- Similar discussion: Boyer-Moore with custom predicate

Compile-Time Evaluation

- Classic programming example: factorial function

```
ulong factorial(uint n) {  
    ulong result = 1;  
    foreach (i; 2 .. n) result *= i;  
    return result;  
}  
  
...  
auto f = factorial(10);
```

Compile-Time Evaluation

- Classic *metaprogramming* example: factorial template

```
template factorial(uint n) {  
    static if (n <= 1)  
        enum ulong factorial = 1;  
    else  
        enum ulong factorial =  
            n * factorial(n - 1);  
}  
auto f = factorial!(10);
```

- BTW: `static if` == awesome

Forget What You Just Saw

- If a regular function is evaluated in a compile-time context, the compiler attempts immediate interpretation:

```
// evaluate at runtime  
auto f1 = factorial(10);  
// evaluate at compile time  
enum f2 = factorial(10);  
// only one function for both!  
ulong factorial(uint n) {  
    ... as above ...  
}
```

- Doing compile-time tricks without even learning them.

Operator overloading

Operator overloading

- Essentially just syntactic sugar
- Many languages chose to just avoid it
- Yet very useful

// Harmonic mean of three numbers

```
m = 3 / (1/x + 1/y + 1/z);
```

// versus:

```
m = divide(3, add(add(divide(1, x),  
    divide(1, y)), divide(1, z)));
```


Connection with Language Design

- Conventional wisdom:
 - Increases language size
 - Limited applicability
 - Open to abuse
- However, consider built-in types
 - High demand for convenient syntax
 - Array, associative array indexing
 - Array, string concatenation
 - Arithmetic and logic operations
- Operator overloading might *decrease* the language's size!

To wit: reducing D's Size

- Associative arrays were built in, now library
- Built-in complex types are being replaced by library types
- Built-in arrays are being replaced by library types

Issues in Defining Operator Overloading

- Definition syntax
- Semantics
- Efficiency
- Code proliferation
 - e.g., exploiting redundancy
- Teaching and understanding
 - Impact on the rest of the language

D's Approach: Lowering

- All uses of operators for user-defined types are lowered to regular method calls
- Example: `++a` rewritten as `a.opUnary! "++" ()`
- Keeps core language small
- Understanding semantics is easy
- Name of operator passed during compilation
- Same method called for different operators
 - Why is this useful?

Binary Operators

- $a + b \rightarrow a.\text{opBinary! "+" (b) and } b.\text{opBinaryRight! "+" (a)}$
- Neither works: “Operator not defined”
- Both work: “Ambiguous operator”
- One works: realize rewrite

mixin For The Win

- Many times operators can be grouped

```
struct CheckedInt(N) if (isIntegral!N) {
    private N value;
    this(N value) {
        this.value = value;
    }
    CheckedInt opUnary(string op)()
        if (op == "+" || op == "-" || op == "~") {
            return CheckedInt(mixin(op ~ "value"));
        }
    ...
}
```

mixin For More Win

```
struct CheckedInt(N) if (isIntegral!N) {
    ...
    // Shift
    CheckedInt opBinary(string op)(CheckedInt rhs)
        if (op == "<<" || op == ">>" || op == ">>>") {
            enforce(rhs.value >= 0 && rhs.value <= N.sizeof * 8);
            return CheckedInt(mixin("value" ~ op ~ "rhs.value"));
        }
    // Logic (no checking needed)
    CheckedInt opBinary(string op)(CheckedInt rhs)
        if (op == "&" || op == "|" || op == "^") {
            return CheckedInt(mixin("value" ~ op ~ "rhs.value"));
        }
}
```

Comparison

- $a == b \rightarrow$
 - `object.opEquals(a, b)` (classes)
 - `a.opEquals(b)` and `b.opEquals(a)` (structs)
- $a < b \rightarrow a.opCmp(b) < 0$ and $0 < b.opCmp(a)$
 - Only one must work

- Redundancies absorbed by definition
- Careful semantics for classes

Indexing: Extra Care

- Distinguish between
 - read: `c = a[b]`
 - write: `a[b] = c`
 - read-modify-write: `a[b] += c`
- Avoid C++'s mistake
 - `operator[]` can't distinguish `c=a[b]` from `a[b]=c` from `a[b]+=c`
 - Impairs expressiveness (see `std::string`)

Reading

- $c = a[b_1, b_2, \dots, b_k]$

→

$$c = a.opIndex(b_1, b_2, \dots, b_k)$$

Writing

- $a[b_1, b_2, \dots, b_k] = c$

→

`a.opIndexAssign(c, b1, b2, ..., bk)`

Read-Modify-Write

- $a[b_1, b_2, \dots, b_k] += c$

→

$c = a.opIndexOpAssign! "+"(c, b_1, b_2, \dots, b_k)$

- $++a[b_1, b_2, \dots, b_k]$

→

$a.opIndexUnary! "++"(b_1, b_2, \dots, b_k)$

Conclusion

Conclusion

- Q: System-level language without the agonizing pain?

Conclusion

- Q: System-level language without the agonizing pain?
- Q: Application language without the boredom?

Conclusion

- Q: System-level language without the agonizing pain?
- Q: Application language without the boredom?
- Q: Principled language without the high-brow snobbery?

Conclusion

- Q: System-level language without the agonizing pain?
- Q: Application language without the boredom?
- Q: Principled language without the high-brow snobbery?

Conclusion

- Q: System-level language without the agonizing pain?
- Q: Application language without the boredom?
- Q: Principled language without the high-brow snobbery?

- $D = \sum_i w_i q_i$

Conclusion

- Q: System-level language without the agonizing pain?
- Q: Application language without the boredom?
- Q: Principled language without the high-brow snobbery?

- $D = \sum_i w_i q_i$

Grill the Speaker!

Questions & Comments

More about ranges? • Thought the talk was boring • Intriguing! • He lost me at opIndexOpAssign—lysdexia is killing me • Awesome • Went over my head • No, went under my feet • I want to implement that • Too self-congratulatory • I wonder how I can implement binary search • What's for dinner? • Accent is bothersome • How about Phobos vs. Tango? • Real programmers use C • Must... control... fist... of... death...