

Language Documentation

Ruben Saunders

January 2025

Contents

1	Introduction	2
1.1	Overview	2
1.2	The Compiler	2
1.3	“Hello, World”	2
2	Names & Scope	3
2.1	Variables	3
2.1.1	Shadowing	3
2.1.2	Constants	3
3	Basic Syntax	4
3.1	Expressions & Statements	4
3.1.1	Value Categories	4
3.2	Code Blocks	4
4	Types	5
4.1	Primitive Types	5
4.2	The Unit Type	5
4.3	Pointer Types	6
4.4	Function Types	6
4.5	Casting	6
5	Functions	7
5.1	Calling Functions	7
5.2	Returning a Value	7
5.2.1	Unreachable Code	8
5.3	Function Declaration	8
5.4	Function Overloading	8
5.4.1	Overload Resolution	9
6	Operators	10
6.1	Casting	10
6.2	Address-Of & Dereferencing	10
6.3	Register Lookup	11
7	Control Flow	12
8	Namespaces	12
9	Modules & Imports	12
10	Standard Library	12

1 Introduction

This document will contain information about the Edel language. As an educational language, it is designed to be simple and easy to learn and operate, with the goals of being clear and readable. Each section is written from the perspective of little programming knowledge.

1.1 Overview

As the most wide-spread paradigm, this language is imperative by nature. Its syntax is inspired by C, but borrows features from C++ such as namespaces and operator overloading. However, unlike these two languages, names (except variables) are *forward declared*, meaning they can be used prior to their definition. This foregoes the need for header files of explicit forward declarations as in C and C++, and allows a much freer programming experience similar to modern languages.

1.2 The Compiler

A program consists of one file (the entry file), which is compiled by the compiler to produce an assembly output.

```
1 $ ./compiler <input_file> [-o <output_file>] [flags]
```

Where [flags] are as follows:

- `--ast` - prints the parsed program's AST.
- `--[no-]function-placeholder` - changes behaviour of a function declaration; see subsection 5.3.
- `--[no-]indentation` - adds/removes indentation from generated assembly.

Note that the output file is optional.

1.3 “Hello, World”

As is customary, a quick “hello world” program written in Edel.

```
func main() {  
    print("Hello, world\n");  
}
```

2 Names & Scope

A *name* in Edel refers to an identifier, which includes functions, variables, and namespaces. All names have a *scope*, which is the area in which that name exists/may be referenced. That is, outside its scope, a symbol does not exist. The outermost scope is known as the *global* scope. Names defined here will be available anywhere in the current file. The counterpart is the *local* scope, which refers to the latest scope. New scopes are created by using code blocks, for example, in while loops and functions.

2.1 Variables

Using the common analogy, a variable is like a labeled box which stored data. Each box has a *type* which tells us what it stores. For example, “int 3”. More technically, a variable is a location in memory where data can be stored, with its type determining the size of the variable and how its value can be used.

Variables are defined using the `let` keyword, with the variables type following a colon. For example, to define our “int 3” from earlier,

```
let foo: int = 3;
```

If required, multiple variables can be defined in the same `let` statement.

```
let foo: int = 3, bar: float = 3.14;
```

If the type is omitted, it is deduced from the assigned value. Hence, this can only be done if the variable is assigned.

```
let foo = 3, bar = 3.14;
```

(In this case, both variables adopt the types as previously written.)

2.1.1 Shadowing

Edel allows names to be *shadowed*, meaning a variable may re-use a variable’s name.

In non-local shadowing, the previous definition becomes visible again after the current scope is exited.

```
let foo = 1;
{
    let foo = 2;
    // foo = 2
}
// foo = 1
```

Edel also allows local shadowing. In this case, once shadowed, the previous definition is essentially lost and cannot be referenced.

```
let foo = 1;
// foo = 1
let foo = 2;
// foo = 2
```

2.1.2 Constants

Constants are values that, once defined, cannot be changed. Constants may be defined using the `const` keyword in place of `let`.

```
const pi = 3.14159;
```

In the case of multiple definitions, *all* symbols will be defined as `const`.

3 Basic Syntax

Inspired by C, Edel is not whitespace sensitive, with scopes marked by braces `{}`. Semicolons are not required after every line, namely closing braces, but are required after expressions.

Comments come in two forms: single- and multi-line. Single-line comments `// ...` comment out the remainder of the current line; multi-line comments `/* ... */` comment out anything they enclose.

3.1 Expressions & Statements

Generally, a statement is some code which does not produce a value, whereas an expression does. In Edel, *everything* is an expression, with statements, by default, returning the unit type `()`.

Expressions are expected to be terminated by a semicolon. The exception to this is the final expression in a block; if omitted, the block ‘returns’ the result of this expression.

```
let a = { 1 + 2; }; // a = ()
let b = { 1 + 2 }; // b = 3
let c = { let z = 2 * b; z } + 2; // c = 8
```

Note that only the last expression in a block may have their semicolon omitted. By always terminating expressions, the block will act like a statement.

3.1.1 Value Categories

Each expression, be it an operator with operands, literals, or variable names etc., is characterised by a *type* and a *category*. Types, covered in section 4, give meaning to data and what it means in context, whereas the category tells dictates *how* a value can be used.

Edel has two categories, be them l- and r-values. The former is a value which refers to a memory location, such as variables, while the latter has no identifiable location in memory. The *l* and *r* stand for *left* and *right* because the respective values may appear on the left- or right-hand side of an assignment operator. That is,

```
a = 1; // a (lvalue) = 1 (rvalue)
```

As an rvalue has no identifiable location, it cannot be assigned to, nor can it be accessed via the dot `.` operator. Instances of misuse of an l- or rvalue results in an error “expected l/rvalue, got type”.

```
1 = 2; // error: expected lvalue, got i32
true.a; // error: expected lvalue, got bool

namespace maths {}
maths + 1; // error: expected rvalue, got namespace
```

3.2 Code Blocks

As a whitespace-ignorant language, Edel uses blocks to denote structure and scope. A block is introduced by braces `{}` and contains a section of code. As previously mentioned, a block may be an expression or a statement, hence is permitted anywhere they would be expected.

Blocks in Edel introduce a new lexical scope (see section 2 for scope resolution).

4 Types

Continuing the analogy from variables, a type is a label that tells the program what kind of data a variable holds. Every variable and value has a type; a value without a type is meaningless, as a type gives data a meaning.

4.1 Primitive Types

Primitive types are a subset of those provided by the compiler representing atomic values, such as numbers.

Type	Alias	Description
<code>bool</code>		A truth value – either <code>true</code> or <code>false</code>
<code>u8</code> <code>i8</code>	<code>byte</code>	Unsigned 8-bit integer Signed 8-bit integer
<code>u16</code> <code>i16</code>		Unsigned 16-bit integer Signed 16-bit integer
<code>u32</code> <code>i32</code>	<code>int</code>	Unsigned 32-bit integer Signed 32-bit integer
<code>u64</code> <code>i64</code>	<code>long</code>	Unsigned 64-bit integer Signed 64-bit integer
<code>f32</code> <code>f64</code>	<code>float</code> <code>double</code>	32-bit floating-point number 64-bit floating-point number

Table 1: Primitive Types

These types have the following type hierarchy:

- `int8 < int16 < int32 < int64`
- `uint8 < uint16 < uint32 < uint64`
- `uintn < int(n + k)`
- `float32 < float64`
- `[u]intn < float64`
- `[u]intn < float32` where $n < 64$.

Note that `bool` is not related to anything. Unlike in other language, `bool` is treated as a distinct type with abstract true/false values, rather than a synonym for integers 1/0, respectively.

4.2 The Unit Type

The unit type, denoted by empty brackets `()`, is a special type meaning *empty*. Representing an object with no value, it is a zero-width type, meaning a symbol of this type occupies no space. For examples of use, a unit is returned from a block which returns no value.

```
let a = {}; // a: ()
```

Note that a unit is neither an r- nor an l-value, so it cannot be used in expressions. That is, the below program errors before even trying to find a candidate for `+`.

```
() + 1; // error: expected rvalue, got ()
```

4.3 Pointer Types

The use of pointers is designed to be minimal in Edel, but they are available nonetheless. Simply put, a pointer stores the *location* (or *address*) of a variable, which can then be used to later reference and access that data. A pointer is indicated by a star ‘*’ preceding the type; that is, type `*T` is telling the compiler *hey, see this location? Data of type T is stored there.*

Subsection 6.2 explains the use of the `&` and `*` unary operators to manipulate pointers.

Pointers do not fit into the type hierarchy, and cannot be implicitly cast. For safety, it is best not to cast pointer types unless one really knows and understands what is going on.

4.4 Function Types

4.5 Casting

Casting refers to the changing of a type, and can be split into two categories: implicit, and explicit. Note that the destination type of either cast **cannot** be zero-sized.

Implicit casting is where a cast is done behind-the-scenes. Type *a* may be implicitly cast (also known as *coerced*) to another type *b* if and only if $a < b$, that is, *a* is a subtype of *b*. In other words, *a* can be coerced into *b* if *b* is sufficient to represent *a* without data loss. For example, `u8` may be coerced into `u64`, or `float` into `double`, but **not** `i32` into `i16`. Implicit casting is done automatically in the following:

```
let a: double = 3.14; // float -> double
```

On the other hand, explicit casting violates these rules, and hence the cast may be *explicit*. Note that not all datatypes may be cast between each-other using explicit casting, but the subtype relationships may be largely violated.

```
let a: u8 = 42; // error!  
let a = 42 as u8; // i32 -> u8
```

5 Functions

Functions are labelled blocks of code which may be called by the programmer. A function may take some input values, known as *arguments*, and may optionally return a value.

An example of a function which takes two integers and returns an integer:

```
func add(a: int, b: int) -> int {
    return a + b;
}
```

That is, a function is defined using the `func` keyword, followed by a name. Parameters in the form `name: type` are listed, comma-separated, between brackets. Note that if no arguments are required, the brackets may be omitted altogether. The function's return type is listed after an arrow; if no return type is needed, this may be omitted. For example,

```
func foo {
    // ...
}

func bar(a: float) { // equivalent to '-> ()'
    // ...
}
```

5.1 Calling Functions

Functions are called by following the function name with brackets `()` (this is the operator name). Any arguments to be supplied to the function are comma-separated inside these brackets, known as an *argument list*. A function with a matching signature is then found (see subsection 5.4 on function overloading), and its return value produced.

```
func add(a: int, b: int) -> int {
    return a + b;
}

let result = add(1, 4); // result: int = 5
```

Note that `()` is an operator like any other. Therefore, if the subject is not a function type, an overload for `operator()` is searched for. The signature will be the subject type followed by each argument's type. For example,

```
5(3.14, true); // operator()(i32, f32, bool)
```

5.2 Returning a Value

The `return` keyword is used to exit/return from the current function. Optionally, the keyword may be followed by an expression, in which case that value will be returned. The type of the returned expression must match, or be a subtype of, the function's return type. Furthermore, if the function returns a non-unit type, a return statement must be present.

```
func foo {
    return 42; // error: cannot convert i32 to ()
}

func bar -> int {
} // error: missing return statement
```

Note that *all* code paths should return, even if you, the programmer, knows that a path is never taken. That is because the compiler cannot currently recognise, for example, always-false conditions.

```
func foo -> int {
    if false {
    } else {
```

```
        return 42;
    }
} // error: missing return statement
```

5.2.1 Unreachable Code

As `return` exits the function, any code after said statement is *unreachable*. This code will never be executed, and a warning will be reported.

```
func foo {
    return;
    a; // warning: unreachable code detected
}
```

The above code would usually result in an error, as symbol ‘a’ is not defined. However, as the code is not reached, no error is raised.

5.3 Function Declaration

A function declaration is one which has no body. For example,

```
func add(a: int, b: int) -> int;
```

As functions in Edel do not require forward-declaration, what does this do? Depending on the compiler flag `--function-placeholder`, it does one of two things.

Enforce Existence The default, `--no-function-placeholder`, a function declaration is used to inform the programmer about the existence of a function. One common use is in auto-generated standard library files, where the function exists but is implemented internally. If the function does not exist, an error will be raised.

```
func imaginary_function(x: int) -> bool; // error: function was declared but
    does not exist
```

Generate Placeholder In this instance, if a function is declared but does not exist, the function will be created to return zero, ‘()’, false, or null, depending on the return type. This is useful when a function is yet to be implemented. For example, the following are identical.

```
func imaginary_function(x: int) -> bool;
// ==
func imaginary_function(x: int) -> bool {
    return false;
}
```

Note that as this generates a definition, it cannot be re-defined.

5.4 Function Overloading

A function in Edel is not only identified by its name, but also its *signature*. The signature of a function is a combination of its name and parameter types. In an Edel program, a function signature must be unique, or the result would be an error. This means that multiple functions with the same name can exist. This is known as function overloading, and a variant of a function with the same name but different parameters is called an *overload*.

```
func add(a: int, b: int) -> int {
    return a + b;
}
func add(a: float, b: float) -> float {
    return a + b;
}
```


In this case, a reference to `add` no longer just refers to one symbol. Instead, it becomes what is known as an *overload set*, and more information is required for a resolution.

```
add; // error: unable to resolve symbol 'add'
```

5.4.1 Overload Resolution

This is the process by which an overload is selected. Firstly, a list of candidates is generated; this would be all functions with a matching name. A candidate is considered a match if the number of arguments match and the type of each parameter matches. Due to subtyping rules, the latter check may lead to the generating of multiple ‘matching’ candidates. In such a case, the candidate with the highest number of identical types wins.

```
func add(a: int, b: int) -> int;
func add(a: long, b: long) -> long;

add(1, 2); // signature is add(int, int)
```

In the above snippet, both candidates match the callee. However, the former wins as `int == int` while `int < long`. If there are still ties at this point, an error is raised as a candidate cannot be decided. In this case, it is best to explicitly cast one or more arguments to reach the desired candidate.

```
func add(a: int, b: long) -> long;
func add(a: long, b: int) -> long;

add(1, 2); // error: unable to resolve
```

6 Operators

Operators are special symbols which combine one or two expressions, known as unary and binary operators, respectively. The table below lists the built-in operators alongside their behaviour and associativity. Operators start from tightest to loosest precedence, with those of the same precedence grouped.

Table 2: Built-In Operators

Operator	Description	Associativity	Overloadable
<code>a.b</code>	Member access	\longrightarrow	No
<code>a()</code>	Function call	\longrightarrow	Yes
<code>(T) a</code> <code>&a</code> <code>*a</code> <code>registerof a</code>	Primitive cast Address-of Dereference Register lookup	\longleftarrow	No
<code>-a</code> <code>a</code> <code>!a</code>	Negation Bitwise NOT Boolean NOT	\longleftarrow	Yes
<code>a as T</code>	Cast	\longleftarrow	No
<code>a * b</code> <code>a / b</code> <code>a % b</code>	Multiplication Division Modulo (remainder)	\longrightarrow	Yes
<code>a + b</code> <code>a - b</code>	Addition Subtraction	\longrightarrow	Yes
<code>a << b</code> <code>a >> b</code>	Bitwise left shift Bitwise right shift	\longrightarrow	Yes
<code>a == b</code> <code>a != b</code> <code>a < b</code> <code>a <= b</code> <code>a > b</code> <code>a >= b</code>	Relational operators	\longrightarrow	Yes
<code>a & b</code>	Bitwise AND	\longrightarrow	Yes
<code>a ^ b</code>	Bitwise XOR	\longrightarrow	Yes
<code>a b</code>	Bitwise OR	\longrightarrow	Yes
<code>a = b</code>	Assignment	\longleftarrow	No

6.1 Casting

The concept of casting is explained in subsection 4.5. Note that the table above presents two casting operators: `(T) a` and `a as T`.

The former is a *primitive* cast, meaning it only supports casting to primitive, atomic types. This cast is direct and is ignorant of the type of `a` and whether the cast is meaningful.

The latter, using the `as` keyword, is a straightforward explicit cast. There are no restrictions on `T` and, as explained in subsection 4.5, this cast still must obey some rules.

```
let a = (int) 3.14; // ok
let b = (*int) 0; // error
let c = 0 as *int; // ok
```

Also note the difference in precedence: the `as` cast is slightly looser than other unary operators.

```
!1 as bool; // same as '(!1) as bool' NOT '!(1 as bool)'
```

6.2 Address-Of & Dereferencing

Both operations involve pointers, with the former generating pointers, and the latter ‘consuming’ them.

The address-of operator, `&`, calculates the memory address of a symbol. That is, if `a: T`, then `&a` is of type `*T`. Dereferencing, then, is address-of's twin; it retrieved the value stored at a pointer. That is, if `p: *T`, then `*a` is of type `T`.

```
let a: int = 5;
let p: *int = &a;
let b: int = *p; // then a == b
```

The dereference operator returns an lvalue, so it may be assigned to. In this case, the memory pointed to by the pointer will be written to. Continuing the example from above,

```
*p = 10; // a = 10
```

However, `b` will still be 5 as it was assigned the dereferenced `p` beforehand.

6.3 Register Lookup

This is a lower-level operator meant for exploration of the processor's registers.

Expecting a symbol argument, the `registerof` operator returns the index (offset) of where the given symbol is located. This allows one to gain insight into the register allocation algorithm.

For example, given the basic program below, one would expect the following.

```
let a = 5;
registerof a; // = 11, corresponding to $r1
```

If the symbol has not been loaded into registers (i.e., it has not been used at all/recently), the operator returns `-1`.

```
let a: int;
registerof a; // = -1
```

7 Control Flow

8 Namespaces

Namespaces allow the programmer to group related functions, variables etc. together in an environment. Not only does this aid in organisation, but also helps to prevent naming collisions.

A namespace is created using the `namespace` keyword, and may contain any structure which is permitted at the global level.

```
namespace maths {  
    func cos(x: double) -> double;  
}
```

To reference a name outside the namespace, its name must be prefixed by the namespace's name.

```
cos(0); // error: unable to resolve symbol 'sin'  
maths.cos(0); // = 1
```

9 Modules & Imports

10 Standard Library