BATTERY – Basic and Algebraic Data Type-Supported Typed Expressions with Recursion and Polymorphism

1 INTRODUCTION

BATTERY is a toy language for the COSE212 course at Korea University. BATTERY stands for **B**asic and **A**lgebraic Data Type-Supported Typed Expressions with **R**ecursion and PolYmorphism, and it supports the following features:

- unit value (())
- number (integer) values (0, 1, -1, 2, -2, 3, -3, ...)
- boolean values (true and false)
- string values ("", "abc", "def", ...) and string concatenation (++)
- arithmetic operators: negation (-), addition (+), subtraction (-), multiplication (*), division (/), and modulo (%)
- **comparison operators**: equality (== and !=) and relational (<, >, <=, and >=)
- **logical operators**: conjunction (&&), disjunction (||), and negation (!)
- sequences (;)
- conditionals (if-else)
- immutable variable definitions (val)
- first-class functions (=>)
- mutually recursive definitions:
 - lazy variable definitions (lazy val)
 - polymorphic recursive functions (def)
 - polymorphic algebraic data types (enum)
- pattern matching (match)
- exit (exit)
- static type checking

This document is the specification of BATTERY. First, Section 2 describes the concrete syntax, and Section 3 describes the abstract syntax with the desugaring rules. Then, Section 4 describes the type system. Finally, Section 5 describes the big-step operational (natural) semantics of BATTERY.

2 CONCRETE SYNTAX

The concrete syntax of BATTERY is written in a variant of the extended Backus–Naur form (EBNF). The notation <nt> denotes a nonterminal, and "t" denotes a terminal. We use ? to denote an optional element and + (or *) to denote one or more (or zero or more) repetitions of the preceding element. The notation +{A} or *{A} denotes the same as + or *, respectively, but the elements are separated by the element A. We use butnot to denote a set difference to exclude some strings from a producible set of strings. We omit some obvious terminals using the ellipsis (...) notation.

```
<digit> ::= "0" | "1" | "2" | ... | "9"
<number> ::= "-"? <digit>+
<alphabet> ::= "A" | "B" | "C" | ... | "Z" | "a" | "b" | "c" | ... | "z"
<idstart> ::= <alphabet> | "_"
<idcont> ::= <alphabet> | "_" | <digit>
<char> ::= /* any character except '"' */
<string> ::= "\"" <char>* "\""
```

```
<keyword> ::= "Boolean" | "Number" | "String" | "Unit"
             | "case" | "def" | "else" | "exit" | "enum" | "false"
            | "if" | "lazy" | "match" | "true" | "val"
          ::= <idstart> <idcont>* butnot <keyword>
<id>
// expressions
<expr> ::= "()" | <number> | "true" | "false" | <string> | <id>
        | "(" <expr> ")" | "{" <expr> "}" | <expr> ";"? <expr>
        | "if" "(" <expr> ")" <expr> "else" <expr>
        | "val" <id> [ ":" <type> ]? "=" <expr> ";"? <expr>
        | <params> "=>" <expr>
        | <expr> [ "[" <type>*{","} "]" ]? "(" <expr>*{","} ")"
        [ <recdef> ";"? ]+ <expr>
        | <expr> "match" "{" [ <mcase> ";"? ]+ "}"
        | "exit" "[" <type> "]" "(" <expr> ")"
// operators
<uop> ::= "-" | "!"
<bop> ::= "+" | "-" | "*" | "/" | "%" | "++" | "&&" | "||"
       | "==" | "!=" | "<" | "<=" | ">" | ">="
// function parameters
<params> ::= "(" <param>*{","} ")"
<param> ::= <id> ":" <type>
// recursive definitions
<recdef> ::= "lazy" "val" <id> "=" <expr>
          | "def" <id> <tvars>? <params> ":" <type> "=" <expr>
           | "enum" <id> <tvars>? "{" [ <variant> ";"? ]+ "}"
// type variables
<tvars> ::= "[" <id>*{","} "]"
// variants
<variant> ::= "case" <id> "(" [ <id> ":" <type> ]*{","} ")"
// match cases
<mcase> ::= "case" <id> "(" <id>*{","} ")" "=>" <expr>
// types
<type> ::= "(" <type> ")" | "Unit" | "Number" | "Boolean" | "String"
        | <id> [ "[" <type>*{","} "]" ]?
        <tvars>? <type> "=>" <type>
         | <tvars>? "(" <type>*{","} ")" "=>" <type>
```

Duplicate field names are not allowed in record expressions and record types. For types, the arrow (=>) operator is right-associative. For expressions, the precedence and associativity of operators are defined as follows:

Description	Operator	Precedence	Associativity
Unary	-, !	8	right
Multiplicative	*, /, %	7	
Additive	++, +, -	6	
Relational	<, <=, >, >=	5	
Equality	==, !=	4	left
Logical Conjunction	&&	3	
Logical Disjunction	11	2	
Pattern Matching	match	1	

3 ABSTRACT SYNTAX

The abstract syntax of BATTERY is defined as follows:

```
Expressions \mathbb{E} \ni e := ()
                                          (EUnit)
                                                                    |e == e
                                                                                                                  (EEq)
                                          (ENum)
                                                                    |e| < e
                                                                                                                  (ELt)
                             |b|
                                          (EBool)
                                                                    | e; e
                                                                                                                  (ESeq)
                                                                    | if (e) e else e
                             | s
                                          (EStr)
                                                                                                                  (EIf)
                                                                    | val x [:\tau]^? = e; e
                             |x|
                                          (EId)
                                                                                                                  (EVal)
                                                                    |\lambda([x:\tau]^*).e
                                          (EAdd)
                                                                                                                  (EFun)
                                                                    |e[\tau^*](e^*)
                             | e * e
                                          (EMul)
                                                                                                                  (EApp)
                                                                    |d^+e|
                                                                                                                  (ERecDefs)
                             | e / e
                                          (EDiv)
                             | e % e
                                          (EMod)
                                                                    | e \text{ match } \{ [case \ x(x^*) \Rightarrow e]^+ \}
                                                                                                                  (EMatch)
                             | e ++ e
                                          (EConcat)
                                                                    |\operatorname{exit}[\tau](e)|
                                                                                                                  (EExit)
         Recursive Definitions \mathbb{D} \ni d ::= 1 \text{ azy } x : \tau = e
                                                                                                     (LazyVal)
                                                    | \det x[\alpha^*]([x:\tau]^*):\tau = e
                                                                                                     (RecFun)
                                                    |\operatorname{enum} t[\alpha^*] \{ [\operatorname{case} x([x:\tau]^*)]^+ \}  (TypeDef)
    Types \mathbb{T} \ni \tau ::= \text{unit}
                                                (UnitT)
                                                                   Identifiers
                                                                                            x \in \mathbb{X}
                                                                                                           (String)
                         num
                                                (NumT)
                                                                   Numbers
                                                                                            n \in \mathbb{Z}
                                                                                                           (BigInt)
                                                                                            b \in \mathbb{B}
                         bool
                                                (BoolT)
                                                                   Booleans
                                                                                                           (Boolean)
                                                                   Strings
                                                                                            s \in \mathbb{S}
                         str
                                                (StrT)
                                                                                                           (String)
                                                                                            t \in \mathbb{X}_t
                         |t[\tau^*]
                                                (IdT)
                                                                   Type Names
                                                                                                           (String)
                                                (IdT)
                                                                   Type Variables
                                                                                            \alpha \in \mathbb{X}_{\alpha}
                                                                                                           (String)
                         \mid [\alpha^*](\tau^*) \to \tau \quad (ArrowT)
```

For type names $t[\tau^*]$ and arrow types $[\alpha^*](\tau^*) \to \tau$, we omit the square brackets (t and $(\tau^*) \to \tau$) when their type arguments (τ^*) or type variables (α^*) are empty, respectively. The types or semantics of the remaining cases are defined with the following desugaring rules:

```
 \begin{split} \mathcal{D} \llbracket - e \rrbracket &= \mathcal{D} \llbracket e \rrbracket * (-1) \\ \mathcal{D} \llbracket ! \ e \rrbracket &= \mathrm{if} \ (\mathcal{D} \llbracket e \rrbracket) \ \mathrm{false} \ \mathrm{else} \ \mathrm{true} \\ \mathcal{D} \llbracket e_1 - e_2 \rrbracket &= \mathcal{D} \llbracket e_1 \rrbracket + \mathcal{D} \llbracket - e_2 \rrbracket \\ \mathcal{D} \llbracket e_1 \ \mathrm{\&\&e_2} \rrbracket = \mathrm{if} \ (\mathcal{D} \llbracket e_1 \rrbracket) \ \mathcal{D} \llbracket e_2 \rrbracket \ \mathrm{else} \ \mathrm{false} \\ \mathcal{D} \llbracket e_1 \mid | \ e_2 \rrbracket = \mathrm{if} \ (\mathcal{D} \llbracket e_1 \rrbracket) \ \mathrm{true} \ \mathrm{else} \ \mathcal{D} \llbracket e_2 \rrbracket \\ \end{split}
```

The omitted cases recursively apply the desugaring rule to sub-expressions.

4 TYPE SYSTEM

This section explains type system of BATTERY, and we use the following notations:

Type Environments
$$\Gamma \in (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}) \times (\mathbb{X}_t \xrightarrow{\text{fin}} (\mathbb{X}_{\alpha}^* \times (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^*))) \times \mathcal{P}(\mathbb{X}_{\alpha})$$
 (TypeEnv)

A type environment Γ consists of three components: 1) a variable mapping $\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}$ that maps variables to their types, 2) a type name mapping $\mathbb{X}_t \xrightarrow{\text{fin}} (\mathbb{X}_{\alpha}^* \times (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^*))$ that maps type names to their type variables and commutative variants, and 3) a set of type variables that are currently in scope. In the type system, type checking is defined with the following typing rules:

$$\tau \text{-EUnit} \ \, \frac{\Gamma \vdash e : \tau}{\Gamma \vdash () : \text{unit}} \ \, \tau \text{-ENum} \ \, \frac{\tau \text{-ENum}}{\Gamma \vdash n : \text{num}}$$

$$\tau \text{-EBool} \ \, \frac{\tau \text{-Ebool}}{\Gamma \vdash b : \text{bool}} \ \, \tau \text{-EStr} \ \, \frac{\tau \text{-ENum}}{\Gamma \vdash s : \text{str}} \ \, \tau \text{-EId} \ \, \frac{x \in \text{Domain}(\Gamma)}{\Gamma \vdash x : \Gamma(x)}$$

$$\tau \text{-EAdd} \ \, \frac{\Gamma \vdash e_1 : \text{num}}{\Gamma \vdash e_1 + e_2 : \text{num}} \ \, \Gamma \text{-Equil} \ \, \frac{\Gamma \vdash e_1 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}} \ \, \frac{\Gamma \vdash e_2 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}}$$

$$\tau \text{-EMod} \ \, \frac{\Gamma \vdash e_1 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}} \ \, \frac{\Gamma \vdash e_2 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}}$$

$$\tau \text{-EConcat} \ \, \frac{\Gamma \vdash e_1 : \text{str}}{\Gamma \vdash e_1 : \text{eq} : \text{str}} \ \, \frac{\Gamma \vdash e_2 : \text{str}}{\Gamma \vdash e_1 : \text{eq} : \text{str}}$$

$$\tau \text{-ELt} \ \, \frac{\Gamma \vdash e_1 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}} \ \, \frac{\Gamma \vdash e_2 : \text{num}}{\Gamma \vdash e_1 : \text{eq} : \text{num}}$$

$$\tau \text{-ESeq} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash e_1 : \text{eq} : \tau_2} \ \, \tau \text{-EIf} \ \, \frac{\Gamma \vdash e_0 : \text{bool}}{\Gamma \vdash e_1 : \text{eq} : \text{str}}$$

$$\tau \text{-EVal} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash e_1 : \text{eq} : \text{eq} : \text{eq}} \ \, \tau \text{-EVal}_T \ \, \frac{\Gamma \vdash e_0 : \text{bool}}{\Gamma \vdash e_1 : \tau_1} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq} : \text{eq} : \text{eq}}$$

$$\tau \text{-EVal} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq} : \text{eq} : \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq} : \text{eq}} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq}}$$

$$\tau \text{-EVal} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq} : \text{eq}} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq}} \ \, \tau \text{-EVal}_T \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_2 : \tau_2}{\Gamma \vdash \text{eq}}$$

$$\tau \text{-EVal} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash \text{eq}} \ \, \frac{\Gamma \vdash e_1 : \tau_1}{\Gamma \vdash$$

$$\tau - \mathsf{ERecDefs} \ \frac{\Gamma \vdash d_1 \vdash \Gamma_1 \quad \dots \quad \Gamma_{n-1} \vdash d_n \vdash \Gamma_n \qquad \Gamma_n \vDash d_1 \quad \dots \quad \Gamma_n \vDash d_n \qquad \Gamma_n \vdash e : \tau \qquad \Gamma \vdash \tau}{\Gamma \vdash d_1; \dots; d_n; e : \tau}$$

$$\begin{split} \Gamma \vdash e : t\big[\tau_1, \dots, \tau_m\big] & \Gamma(t) = \big[\alpha_1, \dots, \alpha_m\big]\big\{x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})\big\} \\ \forall 1 \leq i \leq n. \ \Gamma_i = \Gamma\big[x_{i,1} : \tau_{i,1}\big[\alpha_1 \leftarrow \tau_1, \dots, \alpha_m \leftarrow \tau_m\big], \dots, x_{i,m_i} : \tau_{i,m_i}\big[\alpha_1 \leftarrow \tau_1, \dots, \alpha_m \leftarrow \tau_m\big]\big] \\ \hline \tau \vdash \text{EMatch} & \frac{\Gamma_1 \vdash e_1 : \tau_1 & \dots & \Gamma_n \vdash e_n : \tau_n & \tau_1 \equiv \dots \equiv \tau_n}{\Gamma \vdash e \text{ match } \big\{ \text{ case } x_1(x_{1,1}, \dots, x_{1,m_1}) \Rightarrow e_1; \ \dots; \ \text{ case } x_n(x_{n,1}, \dots, x_{n,m_n}) \Rightarrow e_n \ \big\} : \tau_1 \end{split}$$

$$\tau \text{-EExit} \ \frac{\Gamma \vdash \tau \qquad \Gamma \vdash e : \mathsf{str}}{\Gamma \vdash \mathsf{exit}[\tau](e) : \tau}$$

the following type environment update rules for recursive definitions:

the following typing rules for recursive definitions:

$$\Gamma \models d$$

$$\text{LazyVal} \ \frac{\Gamma \vdash \tau_0 \qquad \Gamma \vdash e_1 : \tau_1 \qquad \tau_0 \equiv \tau_1}{\Gamma \models \text{lazy} \ x \colon \tau_0 = e_1}$$

$$\text{RecFun} \ \frac{\alpha_1 \notin \text{Domain}(\Gamma) \qquad \ldots \qquad \alpha_m \notin \text{Domain}(\Gamma) \qquad \Gamma' = \Gamma[\alpha_1, \ldots, \alpha_m]}{\Gamma' \vdash \tau_1 \qquad \ldots \qquad \Gamma' \vdash \tau \qquad \Gamma' \vdash \tau_1 \quad \ldots, \tau_n \colon \tau_n] \vdash e \colon \tau' \qquad \tau \equiv \tau'}{\Gamma \models \text{def} \ x[\alpha_1, \ldots, \alpha_m] \ (x_1 \colon \tau_1, \ldots, x_n \colon \tau_n) \colon \tau = e}$$

$$\text{TypeDef} \ \frac{\alpha_1 \notin \text{Domain}(\Gamma) \qquad \ldots \qquad \alpha_m \notin \text{Domain}(\Gamma)}{\Gamma' = \Gamma[\alpha_1, \ldots, \alpha_m] \qquad \Gamma' \vdash \tau_{1,1} \qquad \ldots \qquad \Gamma' \vdash \tau_{n,m_n}}{\Gamma' \vdash \tau_{1,1} \qquad \ldots \qquad \Gamma' \vdash \tau_{n,m_n}}$$

$$\Gamma \models \text{enum} \ t[\alpha_1, \ldots, \alpha_m] \ \begin{cases} \text{case} \ x_1(x_{1,1} \colon \tau_{1,1}, \ldots, x_{1,m_1} \colon \tau_{1,m_1}); \\ \ldots \\ \text{case} \ x_n(x_{n,1} \colon \tau_{n,1}, \ldots, x_{n,m_n} \colon \tau_{n,m_n}) \end{cases}$$

the following rules for well-formedness of types:

and the following rules for type equivalence:

5 SEMANTICS

We use the following notations in the semantics:

Environments
$$\sigma \in \mathbb{X} \xrightarrow{\text{fin}} \mathbb{V}$$
 (Env)

The big-step operational (natural) semantics of BATTERY is defined as follows:

with the following environment update rules for recursive definitions:

the following auxiliary function:

$$\begin{array}{c} & \boxed{\operatorname{eq}: \mathbb{V} \times \mathbb{V} \to \mathbb{B}} \\ \operatorname{eq}((),()) = \operatorname{true} & \operatorname{eq}(n,n') = (n=n') & \operatorname{eq}(b,b') = (b=b') & \operatorname{eq}(s,s') = (s=s') \\ \operatorname{eq}(x(v_1,\ldots,v_n),x'(v_1',\ldots,v_n')) = (x=x') \wedge \operatorname{eq}(v_1,v_1') \wedge \ldots \wedge \operatorname{eq}(v_n,v_n') \\ \operatorname{eq}(_,_) = \operatorname{false}(\operatorname{otherwise}) \end{array}$$