

Correct-by-Construction Pretty-Printing

Nils Anders Danielsson

IFIP WG 2.1 Meeting #70,
Schloss Reisenburg, 2013-07-04



Pretty-printing

```
add (mul 1 2) (mul 3 (add 4 5))
```



```
1 * 2 + 3 * (4 + 5)
```

```
1 * 2 +  
3 * (4 + 5)
```

Classical pretty-printing combinators

Implemented by user:

$$\textit{pretty} : A \rightarrow \textit{Doc}$$

Combinator interface:

$$\textit{Doc} : \textit{Set}$$
$$\textit{render} : \mathbb{N} \rightarrow \textit{Doc} \rightarrow \textit{String}$$
$$\textit{text} : \textit{String} \rightarrow \textit{Doc}$$
$$_ \diamond _ : \textit{Doc} \rightarrow \textit{Doc} \rightarrow \textit{Doc}$$
$$\textit{line} : \textit{Doc}$$
$$\vdots$$

Correctness

Assume that we have a parser:

$$\textit{parse} : \textit{String} \rightarrow \textit{List } A$$

Round-tripping property:

$$\forall w x \rightarrow x \in \textit{parse} (\textit{render } w (\textit{pretty } x))$$

How can this property be proved?

This work

I use types to ensure that the round-tripping property holds by construction.

Documents

Before:

$$Doc : Set$$

Now:

$$Grammar : Set \rightarrow Set$$
$$Doc : Grammar A \rightarrow A \rightarrow Set$$

Grammars

Relational semantics for grammars:

$$_ \in _ \cdot _ : A \rightarrow \textit{Grammar} \ A \rightarrow \textit{String} \rightarrow \textit{Set}$$

$x \in g \cdot s$ means that the string s and value x are generated by g .

Pretty-printers

Before:

$$\textit{pretty} : A \rightarrow \textit{Doc}$$

Now:

$$\begin{aligned} g & : \textit{Grammar } A \\ \textit{pretty} & : (x : A) \rightarrow \textit{Doc } g \ x \end{aligned}$$

Renderers

Before:

$$\text{render} : \mathbb{N} \rightarrow \text{Doc} \rightarrow \text{String}$$

Now:

$$\begin{aligned} \text{render} & : \mathbb{N} \rightarrow \text{Doc } g \ x \rightarrow \text{String} \\ \text{parsable} & : \forall w (d : \text{Doc } g \ x) \rightarrow \\ & \quad x \in g \cdot \text{render } w \ d \end{aligned}$$

Correctness (by construction):

$$\forall w \ x \rightarrow x \in g \cdot \text{render } w \ (\text{pretty } x)$$

No guarantee of “prettiness”.

Overview

In talk:

- ▶ Grammars.
- ▶ Documents.
- ▶ Simple examples.

Not in talk:

- ▶ Renderer (based on Wadler's).

Grammars

For simplicity: regular expressions.

data *Grammar* : *Set* \rightarrow *Set*₁ **where**

\emptyset : *Grammar* *A*

ε : *A* \rightarrow *Grammar* *A*

char : *Char* \rightarrow *Grammar* *Char*

$_ \circledast _$: *Grammar* (*A* \rightarrow *B*) \rightarrow *Grammar* *A* \rightarrow
Grammar *B*

$_ | _$: *Grammar* *A* \rightarrow *Grammar* *A* \rightarrow
Grammar *A*

$_ \star _$: *Grammar* *A* \rightarrow *Grammar* (*List* *A*)

Semantics of grammars

$$\overline{x \in \varepsilon x \cdot []} \qquad \overline{c \in \text{char } c \cdot [c]}$$

$$\frac{f \in g_1 \cdot s_1 \quad x \in g_2 \cdot s_2}{f x \in g_1 \otimes g_2 \cdot s_1 \uplus s_2}$$

$$\frac{x \in g_1 \cdot s}{x \in g_1 \mid g_2 \cdot s} \qquad \frac{x \in g_2 \cdot s}{x \in g_1 \mid g_2 \cdot s}$$

$$\frac{xs \in \varepsilon [] \mid \varepsilon _ :: _ \otimes g \otimes g \star \cdot s}{xs \in g \star \cdot s}$$

Semantics of grammars

$$\overline{x \in \varepsilon x \cdot []}$$

$$\overline{c \in \text{char } c \cdot [c]}$$

$$\frac{f \in g_1 \cdot s_1 \quad x \in g_2 \cdot s_2}{f x \in g_1 \otimes g_2 \cdot s_1 \uplus s_2}$$

$$\frac{x \in g_1 \cdot s}{x \in g_1 \mid g_2 \cdot s}$$

$$\frac{x \in g_2 \cdot s}{x \in g_1 \mid g_2 \cdot s}$$

$$\overline{[] \in g \star \cdot []}$$

$$\frac{x \in g \cdot s_1 \quad xs \in g \star \cdot s_2}{x :: xs \in g \star \cdot s_1 \uplus s_2}$$

Some grammar combinators

$_ \triangleleft^* _ : \text{Grammar } A \rightarrow \text{Grammar } B \rightarrow$
 $\text{Grammar } A$

$g_1 \triangleleft^* g_2 = \varepsilon (\lambda x _ \rightarrow x) \circledast g_1 \circledast g_2$

$_ + : \text{Grammar } A \rightarrow \text{Grammar } (\text{List } A)$

$g + = \varepsilon _ :: _ \circledast g \circledast g \star$

Some grammar combinators

whitespace : Grammar Char

whitespace = char ' ' | char '\n'

string : String → Grammar String

string [] = ε []

string (c :: s) = ε _::_ * char c * *string* s

Documents

data $Doc : Grammar A \rightarrow A \rightarrow Set_1$ **where**

- $_{-} \diamond _{-}$: $Doc\ g_1\ f \rightarrow Doc\ g_2\ x \rightarrow Doc\ (g_1 \otimes g_2)\ (f\ x)$
- text** : $Doc\ (string\ s)\ s$
- line** : $Doc\ (\epsilon\ unit \leftarrow \otimes\ whitespace\ +)\ unit$
- nest** : $\mathbb{N} \rightarrow Doc\ g\ x \rightarrow Doc\ g\ x$
- group** : $Doc\ g\ x \rightarrow Doc\ g\ x$
- embed** : $(\forall\ s \rightarrow x_1 \in g_1 \cdot s \rightarrow x_2 \in g_2 \cdot s) \rightarrow Doc\ g_1\ x_1 \rightarrow Doc\ g_2\ x_2$

Defined document combinators

To handle $-|-$:

left : $Doc\ g_1\ x \rightarrow Doc\ (g_1\ |\ g_2)\ x$

left $d = \text{embed} \dots d$

right : $Doc\ g_2\ x \rightarrow Doc\ (g_1\ |\ g_2)\ x$

right $d = \text{embed} \dots d$

Embedding proofs:

$\forall s \rightarrow x \in g_1 \cdot s \rightarrow x \in g_1\ |\ g_2 \cdot s$

$\forall s \rightarrow x \in g_2 \cdot s \rightarrow x \in g_1\ |\ g_2 \cdot s$

Defined document combinators

To handle ε :

$empty : Doc (\varepsilon x) x$

$empty = \text{embed} \dots (\text{text} \{s = ""\})$

Embedding proof:

$$\begin{array}{l} \forall s \rightarrow "" \in string "" \cdot s \rightarrow \\ \quad x \in \varepsilon x \quad \cdot s \end{array}$$

Defined document combinators

To handle $_{\langle * }_{\rangle}$:

$$\begin{aligned} _ \langle _ _ : \text{Doc } g_1 x \rightarrow \text{Doc } g_2 y \rightarrow \\ \text{Doc } (g_1 \langle * g_2) x \\ d_1 \langle d_2 = \text{empty} \diamond d_1 \diamond d_2 \end{aligned}$$

Recall that $g_1 \langle * g_2 = \varepsilon (\lambda x _ \rightarrow x) \otimes g_1 \otimes g_2$.

Simple example

$bit : Grammar Bool$

$bit = \epsilon \text{ true } \langle * \rangle string "1"$
 $\quad | \epsilon \text{ false } \langle * \rangle string "0"$

$bit_p : (b : Bool) \rightarrow Doc bit b$

$bit_p b = ? \quad -- Doc bit b$

Simple example

bit : Grammar Bool

bit = ϵ true $\langle * \rangle$ string "1"
 | ϵ false $\langle * \rangle$ string "0"

*bit*_p : (b : Bool) → Doc bit b

*bit*_p true = ? -- Doc bit true

*bit*_p false = ? -- Doc bit false

Simple example

$bit : Grammar Bool$

$bit = \epsilon \text{ true} \langle * \rangle string "1"$
 $\quad | \epsilon \text{ false} \langle * \rangle string "0"$

$bit_p : (b : Bool) \rightarrow Doc bit b$

$bit_p \text{ true} = left ? \quad -- Doc (\epsilon \text{ true} \langle * \rangle string "1")$
 $\quad \quad \quad \quad \quad -- \quad \quad \quad \text{true}$

$bit_p \text{ false} = ? \quad \quad -- Doc bit \text{ false}$

Simple example

bit : Grammar Bool

bit = ε true $\langle * \rangle$ string "1"
 | ε false $\langle * \rangle$ string "0"

bit_p : (*b* : Bool) \rightarrow Doc bit *b*

bit_p true = left (? $\langle \diamond$?) -- Doc (ε true) true

-- Doc (string "1") *s*

bit_p false = ?

-- Doc bit false

Simple example

$bit : Grammar Bool$

$bit = \epsilon \text{ true} \triangleleft^* string "1"$

$\quad | \epsilon \text{ false} \triangleleft^* string "0"$

$bit_p : (b : Bool) \rightarrow Doc bit b$

$bit_p \text{ true} = left (empty \triangleleft ?) \quad -- Doc (string "1") s$

$bit_p \text{ false} = ? \quad -- Doc bit \text{ false}$

Simple example

bit : Grammar Bool

bit = ϵ true \triangleleft^* string "1"
 | ϵ false \triangleleft^* string "0"

*bit*_p : (b : Bool) → Doc bit b

*bit*_p true = left (empty \triangleleft text)

*bit*_p false = ?

-- Doc bit false

Simple example

$bit : Grammar Bool$

$bit = \epsilon \text{ true} \triangleleft^* string "1"$
 $\quad | \epsilon \text{ false} \triangleleft^* string "0"$

$bit_p : (b : Bool) \rightarrow Doc bit b$

$bit_p \text{ true} = left (empty \triangleleft \text{ text})$

$bit_p \text{ false} = right (empty \triangleleft \text{ text})$

Simple example

$bit : Grammar Bool$

$bit = \epsilon \text{ true} \triangleleft^* string "1"$
 $\quad | \epsilon \text{ false} \triangleleft^* string "0"$

$bit_p : (b : Bool) \rightarrow Doc bit b$

$bit_p \text{ true} = left (empty \triangleleft \text{ text})$

$bit_p \text{ false} = right (empty \triangleleft \text{ text})$

$render\ 10\ (bit_p\ \text{false}) \equiv "0"$ $\text{false} \in bit \cdot "0"$

$render\ 1\ (bit_p\ \text{true}) \equiv "1"$ $\text{true} \in bit \cdot "1"$

More defined document combinators

To handle $_*$:

$$\begin{aligned} \mathit{nil} &: \mathit{Doc} (g \star) [] \\ \mathit{nil} &= \mathit{embed} \dots \mathit{empty} \end{aligned}$$

Embedding proof:

$$\forall s \rightarrow \begin{array}{l} [] \in \varepsilon [] \cdot s \rightarrow \\ [] \in g \star \cdot s \end{array}$$

More defined document combinators

To handle $_★$:

$$\begin{aligned} \text{cons} &: \text{Doc } g \ x \rightarrow \text{Doc } (g \ \star) \ xs \rightarrow \\ &\quad \text{Doc } (g \ \star) \ (x :: xs) \\ \text{cons } d_1 \ d_2 &= \text{embed } \dots \ (\text{empty} \ \diamond \ d_1 \ \diamond \ d_2) \end{aligned}$$

Embedding proof:

$$\begin{aligned} \forall s \rightarrow x :: xs \in \varepsilon \ _::_ \ \textcircled{*} \ g \ \textcircled{*} \ g \ \star \cdot s &\rightarrow \\ x :: xs \in g \ \star \cdot s & \end{aligned}$$

Swallowing trailing whitespace

$symbol : String \rightarrow Grammar String$
 $symbol\ s = string\ s \triangleleft^* whitespace \star$

$symbol-nil : Doc (symbol\ s)\ s$
 $symbol-nil = \text{text} \triangleleft nil$

$symbol-line : Doc (symbol\ s)\ s$
 $symbol-line = \text{embed} \dots (\text{text} \triangleleft line)$

Embedding proof:

$\forall s' \rightarrow$
 $s \in string\ s \triangleleft^* (\epsilon\ \text{unit} \triangleleft^* whitespace\ +) \cdot s' \rightarrow$
 $s \in string\ s \triangleleft^* whitespace \star \cdot s'$

Pattern

- ▶ For (almost) every grammar combinator:
one or more document combinators.
- ▶ Embedding proofs in reusable combinators,
ideally not in pretty-printers.

Another example

(Based on an example due to Doaitse and Olaf.)

bit-list : Grammar (List Bool)

bit-list = (bit $\langle \ast$ symbol ";") \star

Another example

bit-list : Grammar (*List Bool*)

bit-list = (*bit* $\langle * \rangle$ *symbol* ";") ★

"1; 0; 0;"

"1; 0;0;\n "

"1;\n\n\n \n 0; 0;"

Another example

bit-list : Grammar (List Bool)

bit-list = (bit $\langle \ast \rangle$ symbol ";") \star

*bit-list*_p : (bs : List Bool) → Doc *bit-list* bs

*bit-list*_p [] = nil

*bit-list*_p (b :: []) = cons (bit_p b $\langle \diamond \rangle$ symbol-nil) nil

*bit-list*_p (b :: bs) =

cons (bit_p b $\langle \diamond \rangle$ group (nest 1 symbol-line))

(*bit-list*_p bs)

Another example

$bit\text{-}list_p : (bs : List Bool) \rightarrow Doc\ bit\text{-}list\ bs$

$bit\text{-}list_p [] = nil$

$bit\text{-}list_p (b :: []) = cons (bit_p b \triangleleft symbol\text{-}nil) nil$

$bit\text{-}list_p (b :: bs) =$
 $cons (bit_p b \triangleleft group (nest\ 1\ symbol\text{-}line))$
 $(bit\text{-}list_p\ bs)$

$bs = true :: false :: true :: true :: false :: []$

$render\ 20\ (bit\text{-}list_p\ bs) \equiv "1; 0; 1; 1; 0;"$

$render\ 10\ (bit\text{-}list_p\ bs) \equiv "1; 0; 1;\n 1; 0;"$

$render\ 6\ (bit\text{-}list_p\ bs) \equiv "1; 0;\n 1; 1;\n 0;"$

Another example

$bit\text{-}list_p : (bs : List Bool) \rightarrow Doc\ bit\text{-}list\ bs$

$bit\text{-}list_p [] = nil$

$bit\text{-}list_p (b :: []) = cons (bit_p b \triangleleft symbol\text{-}nil) nil$

$bit\text{-}list_p (b :: bs) =$
 $cons (bit_p b \triangleleft group (nest\ 1\ symbol\text{-}line))$
 $(bit\text{-}list_p\ bs)$

$bs = true :: false :: true :: true :: false :: []$

$bs \in bit\text{-}list \cdot "1; 0; 1; 1; 0;"$

$bs \in bit\text{-}list \cdot "1; 0; 1;\n 1; 0;"$

$bs \in bit\text{-}list \cdot "1; 0;\n 1; 1;\n 0;"$

More

- ▶ Can use much more general grammar formalism (recursively enumerable languages).
- ▶ More advanced examples available (operators with precedence, an XML-like language).
- ▶ One can prove that the document combinators satisfy certain algebraic properties.

Conclusions

- ▶ Light-weight approach to correct-by-construction pretty-printing.
- ▶ Based on classical pretty-printing, but precisely typed.
- ▶ Separates grammars and pretty-printers.
- ▶ Seems to work well when the pretty-printer follows the grammar's structure.