Adding Apples and Oranges

Martin Erwig
Margaret Burnett

Prevent errors in spreadsheets

An Example

According to Excel:
Aug-00 = 36739
The Problem

Two aspects of spreadsheets:
• up to 40% contain errors
• 55,000,000 end user programmers in 2005 (in the United States)

⇒ Strong need to make spreadsheets more reliable

The Challenge

• Type Systems can be extremely helpful in error detection/prevention, but:
  • An abstract concept of types is very difficult to impart on end users
The Idea

Use vocabulary from the spreadsheet to communicate with the end user

Example:
Column headers

'units'

Adding 'Specific'
Apples and Oranges

Goal: B3+C3 ✔ B3+C4 ERROR!
**Formalization Roadmap**

**Goals:**
1. associate cells with units
2. determine unit-correctness

- Initial unit information is given by headers
- Multiple/nested headers and operations lead to complex units
- Define unit normal form and simplification rules
- Notion of unit-correctness: All cells have units in normal form

---

**Units**

- Values and cells have units
  - `8::Apple`  `B3::Apple`
  - `dependent units`
- Units can be nested
  - `B2::Fruit`  `B3::Fruit[Apple]`
  - `or units`
- Cells might have multiple units
  - `B3::Month[May]&Fruit[Apple]`
  - `and units`
- Operations can generalize units
  - `B3+C3::Apple|Orange`
  - `unit judgments`
  - `B3+C3::Fruit`
  - `And also: units 1 and e`
Headers & Units

- All values in a spreadsheet are units
  - Fruit Apple June 8 50 ...
- A header is a label for a group of cells and defines unit information
  - Fruit: B2 C2
  - Apple: B3 B4
  - May: B3 C3
- Chains of headers define dependent units
  - Fruit[Apple]
  - Month[May]
- A cells can have multiple headers ⇒ & units
  - B3::Month[May]&Fruit[Apple]

Well-formed Units

Unrelated units can be combined with &
- Month[May]&Fruit[Apple]
- Month[May]&Fruit
- Month&Fruit
- Apple&Orange

Units of the same nesting level (>1) that have the same ancestors can be combined with |
- Fruit[Apple]|Fruit[Orange]
- Month[May]|Month[June]
- Month|Fruit
- Fruit[Apple[Green]]|Fruit[Orange]
Unit Inference

1. A cell without a header has unit 1
   - B1::1  D2::1

2. Units propagate through references
   - B3::Fruit[Apple]

3. A cell with a header that contains value V and has unit U has unit U[V]
   - B3::Fruit[Apple]

4. Operations have their own unit transformations
   - \( +_u(U_1, U_2; U) = (U_1|U_2)\&U \)
   - \( \times_u(U_1, 1; U) = U_1\&U \)
   - \( \times_u(1, U_2; U) = U_2\&U \)
   - \( U_2::(\text{Fruit[Apple]}|\text{Fruit[Orange]})\&1 \)

Unit Simplification

\[
\begin{align*}
U_2\&U_1 &= U_2\&U_1 & \text{commutativity} \\
U_2|U_1 &= U_2|U_1 \\
(U_1\&U_2)\&U_3 &= U_1\&(U_2\&U_3) & \text{associativity} \\
(U_1|U_2)|U_3 &= U_1|(U_2|U_3) \\
U\&U &= U \\
U|U &= U & \text{idempotency} \\
U&(U_1|U_2) &= (U\&U_1)|(U\&U_2) & \text{distributivity} \\
1\&U &= U & \text{unit} \\
U[U_1]|U[U_2] &= U[U_1|U_2] & \text{factorization} \\
U[U_1|...|U_k] &= U & \text{generalization(*)} \\
(U_1[U_2])[U_3] &= U_1[U_2[U_3]] & \text{linearization}
\end{align*}
\]
Example

By header (rule 3):
B3::Month[May] & Fruit[Apple]
C3::Month[May] & Fruit[Orange]

By operation + (rule 4):  
D3::Month[May] & Fruit[Apple] |
Month[May] & Fruit[Orange]

Simplification (distribute, factor, generalize):
D3::Month[May] & (Fruit[Apple] | Fruit[Orange])
D3::Month[May] & (Fruit[Apple] | Orange)
D3::Month[May] & Fruit

Two Products

Offline tool to check legacy spreadsheets possible now ✓

Online unit checking integrated into Microsoft Excel
need permission/support from Microsoft...
Future Work

- Units are more fine-grained than types
  ⇒ Define unit-aware semantics to obtain appropriate soundness results

- Header inference
  (1) Predefined unit information
      *May:* Month  *Blue:* Color ...
  (2) Infer headers from table formatting actions
  (3) Spatial analysis
      Infer headers from formula positions

One Conclusion
(from a PADL point of view)

Adding
Functional Programming/Type Systems
and
End-User Programming