

The Weird and Wonderful World of *Excel*

A PL Perspective



File

Home

Insert

Page Layout

Formulas

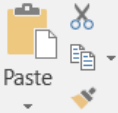
Data

Review

View

Tell me what you want to do

Share



Paste

Clipboard

Calibri 11

B I U

Font



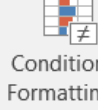
Alignment

Wrap Text

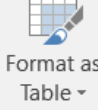
Merge & Center

General

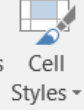
Number



Conditional Formatting



Format as Table



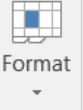
Cell Styles



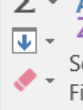
Insert



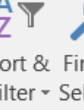
Delete



Format



Sort & Filter



Find & Select



Editing

A1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2																	
3																	
4																	
5																	
6																	
7																	
8																	
9																	
10																	
11																	
12																	
13																	
14																	
15																	
16																	
17																	
18																	
19																	
20																	
21																	
22																	
23																	
24																	
25																	
26																	
27																	
28																	

Sheet1



- 300m active users
- Makes Excel the #1 most used declarative programming language in the world

What?

	A
1	=SIN(A2)
2	=A3+A4
3	2
4	3
5	



$$A1 = \text{SIN}(A2)$$

$$A2 = A3 + A4$$

$$A3 = 2$$

$$A4 = 5$$

Formula Language

Syntax

Recap of Basic Syntax

- Formulas
 - Operations
 - Function Calls
 - Ranges
 - Arrays

Advanced Syntax

- Cross-sheet references (a.k.a 3d references)
- Array-entered formulae (more on this later...)
- Range operations
- Optional[/missing] arguments

Formula Language

Semantics

Type System!

- 4 Basic Types (that can live in cells)
 - Numbers
 - Strings
 - Booleans
 - Errors
- 2 vector types (that cannot live in cells)
 - Ranges
 - Arrays

Types...

- All operations and functions have type signatures
 - $(+) :: \mathbb{N}, \mathbb{N} \rightarrow \mathbb{N}$
- Not all so simple, functions have:
 - Positional arguments
 - Repeated argument group
 - Return type
- Typical:
 - $SUM :: \mathbb{N}; \mathbb{N} \rightarrow \mathbb{N}$ (note the difference with $(+)$)

Types... (cont)

- `"3"+"4"` (?)
 - $(+) :: N, N; \rightarrow N$
 - $"3" :: S, "4" :: S$
- Returns 7 (!)
 - S can be *coerced* to N

Coercions

- N, S, B can be coerced between each other
- Ranges can be coerced to N, S, B
- Arrays do not coerce, they *lift*
 - $\text{SIN}(\{1, 2, 3\}) = \{\text{SIN}(1), \text{SIN}(2), \text{SIN}(3)\}$
 - Multiple unexpected arrays zip:
 - $\{1, 2\} + \{3, 4\} = \{1 + 3, 2 + 4\}$
- Errors cannot be coerced, they *propagate...*

Errors

- **not** “exceptional” values
- Passing an unexpected error to a function will *not* call the function
 - Instead it will immediately return the error unchanged

Ranges

- Many types:

- A1 (cell reference)
 - A1:B2 (2d range)
 - A1 B2 (intersection)
 - (A1, A2) (union)
 - Sheet1:Sheet2!A1:B2 (3d)
- 
- not covered

Ranges... (cont)

- Unexpected ranges:
 - Get *dereferenced*
 - If range is single row/col then pick intersection with home cell
 - Else-if range is on a different sheet, intersect with home cell across sheets
 - Else pick top-left corner

Excel

Implementation

Calc

- “Push” dependency model
 - Dependency graph is forest of DAGs pointing to dependents

	A	B	C
1		=SIN(A2)	
2	1		=COS(B1)
3			
4			

- Any change will propagate forward through the dependencies
- Each DAG can be computed in parallel¹

¹Too complex to talk about here

Array-entered Formulae

- [ctrl]-[shift]-[enter]
 - Inserts a single formula over many cells

	A	B
1	= {1;2}	
2	= {1;2}	
3		
4		

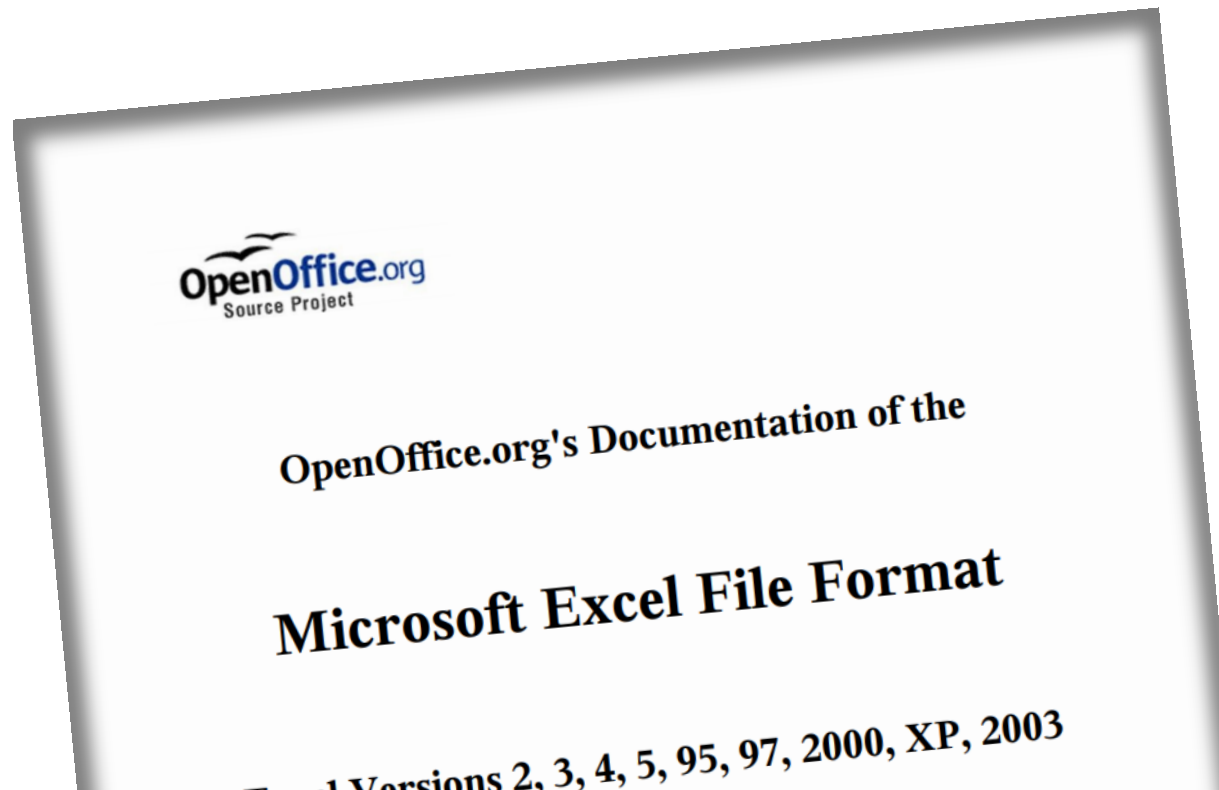
- {A1:A2={1; 2}}

Array-entered... (cont)

- Problem? Overlapping regions
 - {A1:A2=A2:A3+1}
- Excel is fine with this!

Disclaimer

- Source from OpenOffice document:
 - *The Microsoft Excel File Format*



Floats

- Not IEEE-754 compliant
 - No Infinities
 - No NaN
 - No subnormal numbers
- Truncation

Floats... (cont)

- `=SUM(0.1, 0.1, 0.1)`
 - IEEE 754 representation would make this `~0.3000000000444`
 - But `=SUM(0.1, 0.1, 0.1)=0.3` is **TRUE**
- `=(SUM(0.1, 0.1, 0.1) - 0.3) = 0`
 - This is **FALSE**

XLS File Format

- Binary file format
 - Multiple *streams* of data arranged hierarchically
 - Workbook stream
 - Globals stream
 - Worksheet stream
 - Cell stream
 - Compiled formula

Formula Compilation

1. Formula Gets Parsed (this alone is enough for an hours rant)
2. Formula Gets Compiled to Bytecode (!)
3. Bytecode gets interpreted by Excel

Bytecode

- Way too much to cover here
- A small sample of fun looking instructions:
 - PtgAttrGoto
 - PtgAttrSpace
 - PtgAttrSpaceSemi
 - PtgAttrSpaceType
 - PtgInt
 - PtgNum
 - PtgElfRadicalLel (don't ask...)

Bytecode... (cont)

- =1+1
 - Compiles to:

PtgInt 1		PtgInt 1		PtgAdd
0x05	0x01	0x05	0x01	0x03

Bytecode... (cont)

- =1.0+2.0
 - Compiles to:

PtgNum 1.0	PtgNum 1.0	PtgAdd
9 Bytes	9 Bytes	1 Byte

- 17 Bytes!
 - An ASCII encoding would only be 8 ...

Back to floats

- $=(\text{SUM}(0.1, 0.1, 0.1) - 0.3)$

PtgNum 0.1	PtgNum 0.1	PtgNum 0.1	PtgFunc SUM	PtgNum 0.3	PtgSub	PtgParen
9 bytes	9 bytes	9...	3...	9	1	1

- 41 bytes in total!
- If a formula bytecode ends in PtgAdd or PtgSub it truncates the value written to the cell

Parentheses in bytecode?

- =IF(TRUE, 1, 2)
 - Compiles to

Offset	Size	Instruction
0	2	PtgBool TRUE
2	4	PtgAttrIf 11
6	4	PtgAttrSpace 1
10	3	PtgInt 1
13	4	PtgAttrGoto +18
17	4	PtgAttrSpace
21	3	PtgInt 2
24	4	PtgAttrSpace 1
28	4	PtgAttrGoto 3
32	4	PtgFunc IF

End