

Motivation
Survey
What was discussed so far
What areas will be covered next
PURe research team approach
What's next?

Spreadsheet Understanding

A Survey

Alexandra Silva

Departamento de Informática
Universidade do Minho

PURe Workshop, October 2005



Motivation

Survey

What was discussed so far

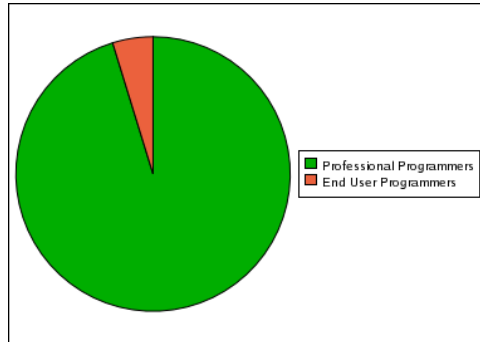
What areas will be covered next

PURe research team approach

What's next?

Big Reasons

- End user programmers



Big Reasons

- End user programmers
- Money – errors in spreadsheets cost millions of dollars to enterprises.



Motivation

Survey

What was discussed so far

What areas will be covered next

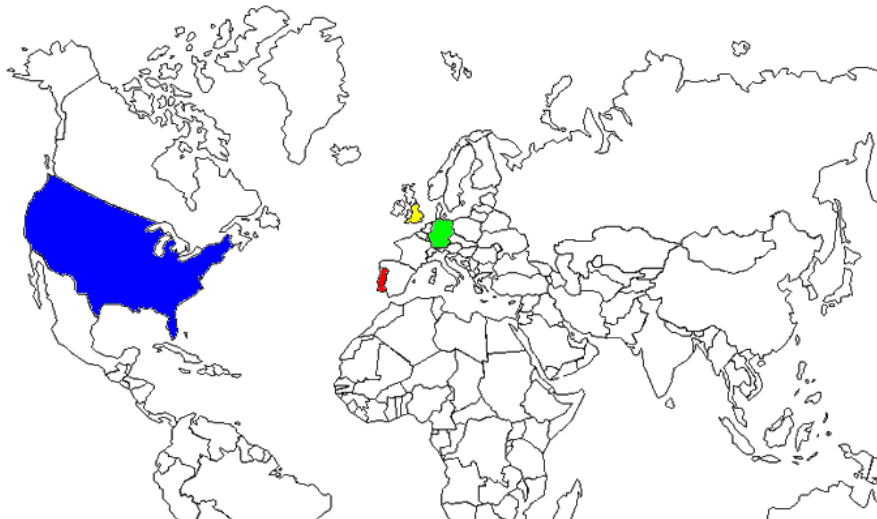
PURe research team approach

What's next?

Martin Erwig *et al.*

Simon Peyton Jones *et al.*

Research all over the world



Alexandra Silva

SpreadSheet Understanding

Motivation

Survey

What was discussed so far
What areas will be covered next
PUR research team approach
What's next?

Martin Erwig *et al.*

Simon Peyton Jones *et al.*

- Martin Erwig and Margaret Burnett



- System of units in spreadsheets;
- Identification of errors based on bad combination of units;
- Automatic generation and maintenance of spreadsheets using *templates*.






What was discussed so far
What areas will be covered next
PURe research team approach
What's next?

Martin Erwig *et al.*

Simon Peyton Jones *et al.*

Adding apples and oranges

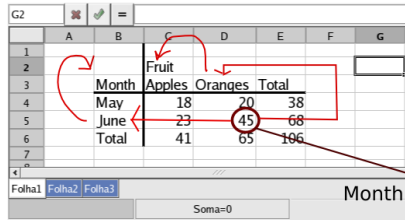
G2	  						
	A	B	C	D	E	F	G
1							
2			Fruit				
3		Month	Apples	Oranges	Total		
4		May	18	20	38		
5		June	23	45	68		
6		Total	41	65	106		
7							
8							
///							
Folha1	Folha2	Folha3	Soma=0				



What was discussed so far
What areas will be covered next
PURe research team approach
What's next?

Martin Erwig *et al.*Simon Peyton Jones *et al.*

Unit system



	A	B	C	D	E	F	G
1							
2			Fruit				
3		Month	Apples	Oranges	Total		
4		May	18	20	38		
5		June	23	45	68		
6		Total	41	65	106		
7							

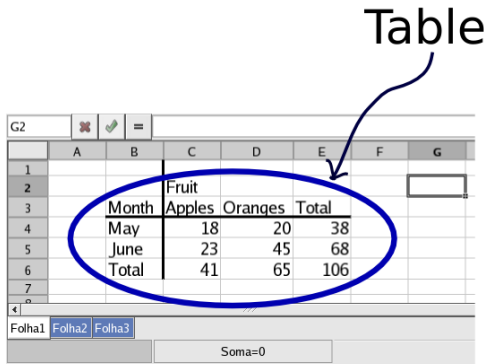
Month[June] & Fruit[Oranges]



Visually Customizing Inference Rules About Apples and Oranges

- Borders identified

Table



	A	B	C	D	E	F	G
1							
2			Fruit				
3		Month	Apples	Oranges	Total		
4		May	18	20	38		
5		June	23	45	68		
6		Total	41	65	106		
7							

Folha1 Folha2 Folha3

Soma=0

What was discussed so far
What areas will be covered next
PURé research team approach
What's next?

Martin Erwig *et al.*Simon Peyton Jones *et al.*

- Groups data

G2							
	A	B	C	D	E	F	G
1							
2			Fruit				
3		Month	Apples	Oranges	Total		
4		May	18	20	38		
5		June	23	45	68		
6		Total	41	65	106		
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
74							
75							
76							
77							
78							
79							
80							
81							
82							
83							
84							
85							
86							
87							
88							
89							
90							
91							
92							
93							
94							
95							
96							
97							
98							
99							
100							

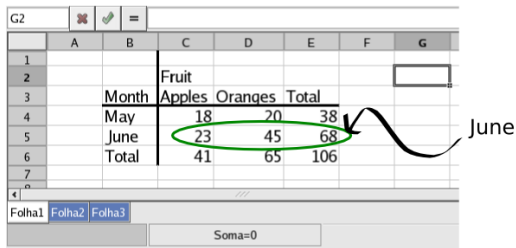
Apples



What was discussed so far
What areas will be covered next
PURe research team approach
What's next?

Martin Erwig *et al.*Simon Peyton Jones *et al.*

- Groups data



	A	B	C	D	E	F	G
1							
2			Fruit				
3		Month	Apples	Oranges	Total		
4		May	18	20	38		
5		June	23	45	68		
6		Total	41	65	106		
7							
8							

Folha1 Folha2 Folha3

Soma=0

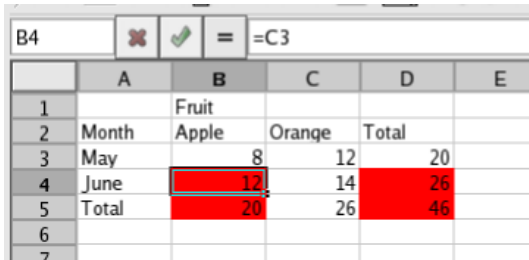


Header and Unit Inference for Spreadsheets Trough Spatial Analysis

- Integrated into Excel
- Identifies reference errors, range errors, etc



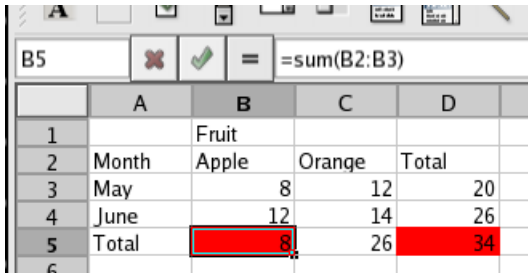
- Identifies **reference errors**, range errors, etc



	A	B	C	D	E
1		Fruit			
2	Month	Apple	Orange	Total	
3	May	8	12	20	
4	June	12	14	26	
5	Total	20	26	46	
6					
7					



- Identifies reference errors, **range errors**, etc



The screenshot shows a spreadsheet application interface. The formula bar at the top displays the formula `=sum(B2:B3)` in cell B5. Below the formula bar is a table with columns A, B, C, and D. The table contains data for months and fruit counts. The cell B5 is highlighted in red, indicating a range error. The error message is "Range error: B2:B3 is not a valid range".

	A	B	C	D
1		Fruit		
2	Month	Apple	Orange	Total
3	May	8	12	20
4	June	12	14	26
5	Total	8	26	34
6				



Motivation

Survey

What was discussed so far

What areas will be covered next

PURe research team approach

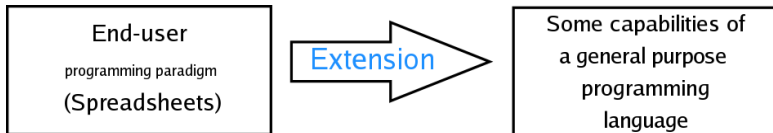
What's next?

Martin Erwig *et al.*

Simon Peyton Jones *et al.*

Simon Peyton Jones *et al.*





- User defined functions
- Matrices



Motivation

Survey

What was discussed so far

What areas will be covered next

PURe research team approach

What's next?

- Unit System
- Identification of tables in spreadsheets (**Ucheck**)
- User defined functions in Excel



Motivation

Survey

What was discussed so far

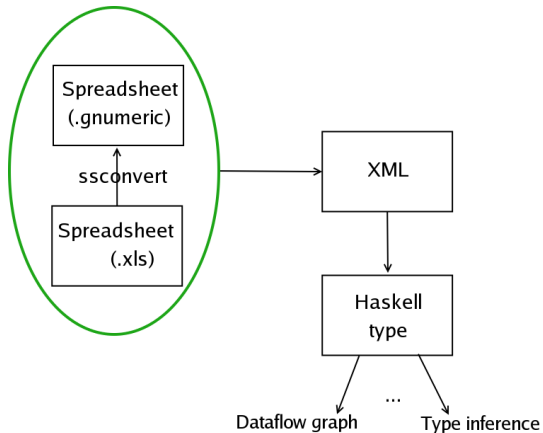
What areas will be covered next

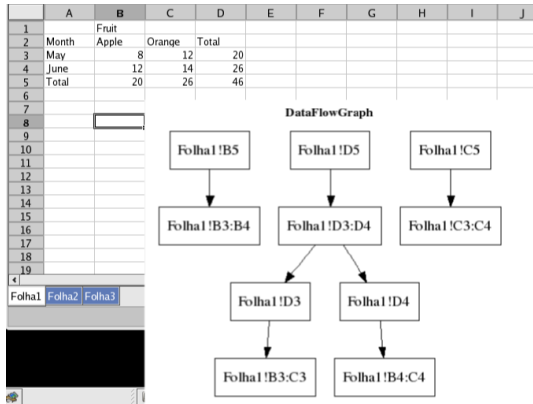
PURe research team approach

What's next?

- Dataflow inside spreadsheets
- Database pointfree theory
- Functional Dependencies







Motivation

Survey

What was discussed so far

What areas will be covered next

PURé research team approach

What's next?

Joost and Cupertino's work

Databases pointfree theory

... and theory becomes simple and concise



Motivation

Survey

What was discussed so far

What areas will be covered next

PURé research team approach

What's next?

Joost and Cupertino's work

Databases pointfree theory

Relational database table \equiv set of n -tuples



Relational database table \equiv set of n -tuples

Number	Name	Mark
35820	"Ana Quintela Castro Ferreira"	18
33567	"David Filipe Oliveira Costa"	17
35672	"Joana Figueiredo Martins"	12
...

\equiv

{ (35820, "Ana Quintela Castro Ferreira", 18), (33567, "David Filipe Oliveira Costa", 17), (35672, "Joana Figueiredo Martins", 12)}



$S = \{\underline{\text{Number}}, \underline{\text{Name}}, \underline{\text{Mark}}\}$ is the *schema* of the table.

Number uniquely determines Name.

Name is **functionally dependent** of Number.



Functional dependency – Traditional definition

Given subsets $x, y \subseteq S$ of the relation scheme S of a relation R , this relation is said to satisfy functional dependency $x \twoheadrightarrow y$ iff all pairs of tuples $t, t' \in R$ which “agree” on x also “agree” on y , that is,

$$\langle \forall t, t' : t, t' \in R : t[x] = t'[x] \Rightarrow t[y] = t'[y] \rangle$$

(Notation $t[x]$ meaning “the values in t of the attributes in x ” will be scrutinized in the sequel)



Functional dependency – Pointfree definition

We say that relation $R : B \leftarrow A$ satisfies the $f \rightarrow g$ functional dependency iff $g \leq f.R^\circ$ or, equivalently, $\ker(f.R^\circ) \subseteq \ker g$.



What's next

- To define n – *ary* tuples (HList)
- Functional dependencies pointfree calculus
- Conceptual analysis(?)



Motivation
Survey
What was discussed so far
What areas will be covered next
PURe research team approach
What's next?

The team



Contributers are welcome:)

Questions?

