

电子表格中表格克隆及相关错误检测

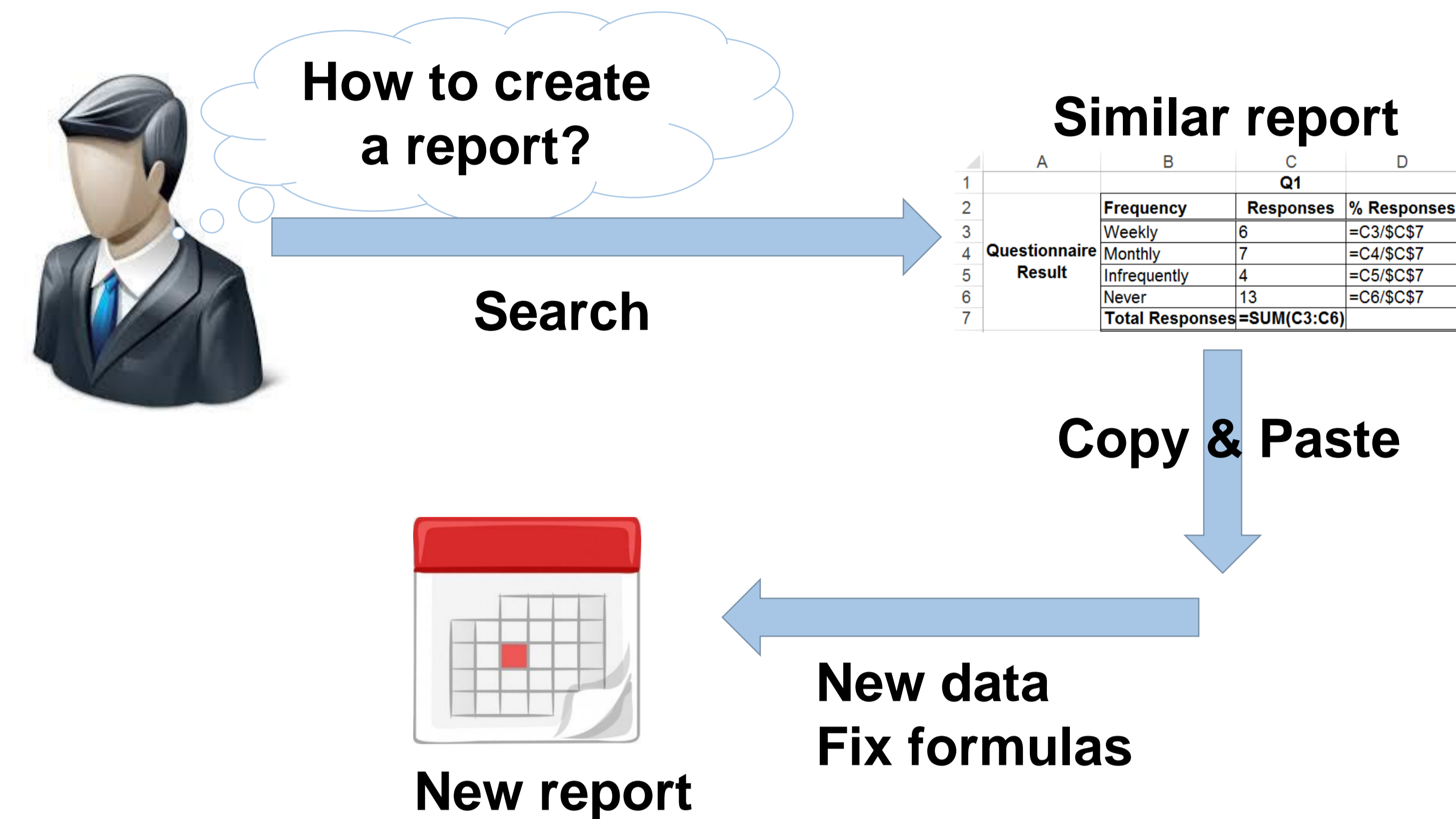
窦文生, 张成志, 高楚舒, 许畅, 徐良, 魏峻

Detecting Table Clones and Smells in Spreadsheets, FSE 2016

联系方式: 窦文生, wensheng@iscas.ac.cn, http://www.tcse.cn/~wsdou

Table Clones and Related Smells

Cloning is common in spreadsheets!



Clone-related smells

	A	B	C
1		Q2	
2	Frequency	Responses	% Responses
3	Weekly	10	=B3/\$B\$7
4	Monthly	4	=B4/\$B\$7
5	Infrequently	4	=B5/\$B\$7
6	Never	12	=B6/\$B\$7
7	Total Responses	=SUM(B3:B6)	

Total responses are \$B\$7

Inconsistency

	A	B	C
1		Q3	
2	Frequency	Responses	% Responses
3	Weekly	3	=B3/30
4	Monthly	2	=B4/30
5	Infrequently	2	=B5/30
6	Never	23	=B6/30
7	Total Responses	30	

Total responses must be 30, and never change!

◆ 21.8% spreadsheets in EUSES contain table clones.

◆ Inconsistencies among table clones indicate smells.

TableCheck

Table clone detection

	A	B	C
1		Q2	
2	Frequency	Responses	% Responses
3	Weekly	10	=B3/\$B\$7
4	Monthly	4	=B4/\$B\$7
5	Infrequently	4	=B5/\$B\$7
6	Never	12	=B6/\$B\$7
7	Total Responses	=SUM(B3:B6)	

Same Headers

	A	B	C
1		Q3	
2	Frequency	Responses	% Responses
3	Weekly	3	=B3/30
4	Monthly	2	=B4/30
5	Infrequently	2	=B5/30
6	Never	23	=B6/30
7	Total Responses	30	

Smell detection

	A	B	C	D
1		Q1		
2	Questionnaire	Frequency	Responses	% Responses
3	Result	Weekly	6	=C3/\$C\$7
4		Monthly	7	=C4/\$C\$7
5		Infrequently	4	=C5/\$C\$7
6		Never	13	=C6/\$C\$7
7	Total Responses	=SUM(C3:C6)		

	A	B	C
1		Q3	
2	Frequency	Responses	% Responses
3	Weekly	3	=B3/30
4	Monthly	2	=B4/30
5	Infrequently	2	=B5/30
6	Never	23	=B6/30
7	Total Responses	30	

Monthly responses / Total (C4/\$C\$7 or B4/\$B\$7)

Monthly responses / 30 (B4/30)

◆ Tables with the same headers are likely to be clones.

◆ As smelly cells normally occur in minority, they can be detected as outliers.

Results

Tool result

	A	B	C
1		Q3	
2	Frequency	Responses	% Responses
3	Weekly	3	=B3/30
4	Monthly	2	=B4/30
5	Infrequently	2	=B5/30
6	Never	23	=B6/30
7	Total Responses	30	

Smelly cells

Referenced Cells

Experimental result on EUSES

- ◆ 14.6% table clones contain smells, and 33.6% smelly cells contain wrong values (harmful).
- ◆ The precision for table clone and smell detection is 92.2% and 85.5%, respectively.
- ◆ Existing approaches can only detect at most 35.6% smells that TableCheck detects.