Fact Checking Quiz: Evaluation
IFIP Working Group 10.4 Summer Meeting
June 22-25, 2017

Saurabh Bagchi
Purdue University

This table shows how the two automated systems – ClaimSpotter and our system Tathya – performed with respect to the 20 statements in the quiz. The objective for the systems is to flag for each statement, if it is check-worthy or not. Further, here, for the check-worthy statements, we indicate how it was rated in the truthometer, the scale being: True, Mostly true, Partly true-partly false, Mostly false, False.

Statement Number	Ground-Truth	ClaimSpotter	TATHYA
1.	Check-worthy (True)	×	✓
2.	Not Check-worthy	✓	×
3.	Check-worthy (Partly False)	×	✓
4.	Not Check-worthy	×	×
5.	Check-worthy (Mostly False)	×	×
6.	Check-worthy (Mostly False)	✓	×
7.	Not Check-worthy	✓	×
8.	Not Check-worthy	✓	×
9.	Check-worthy (Mostly False)	✓	✓
10.	Not Check-worthy	×	×
11.	Check-worthy (True)	×	✓
12.	Not Check-worthy	×	×
13.	Check-worthy (Mostly False)	✓	✓
14.	Not Check-worthy	✓	×
15.	Check-worthy (Mostly False)	×	×
16.	Not Check-worthy	✓	×
17.	Check-worthy (Mostly False)	✓	✓
18.	Check-worthy (Mostly True)	✓	✓
19.	Not Check-worthy	×	×
20.	Not Check-worthy	×	×

We see that TATHYA does much better on flagging check-worthy statements. The not check-worthy ones are from those that TATHYA got wrong (False Positives). Overall TATHYA gains primarily in recall while flagging more false-positives. The false positives are more subjective because non-technical reasons (budgetary reasons, time to publication, etc.) may dictate which statements a journalistic organization decides to check.