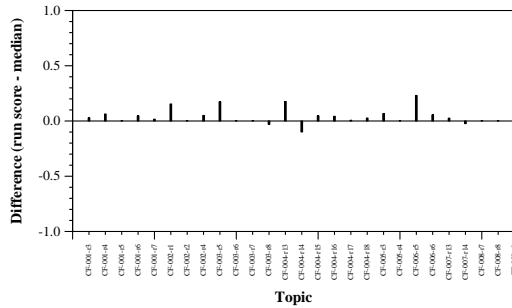


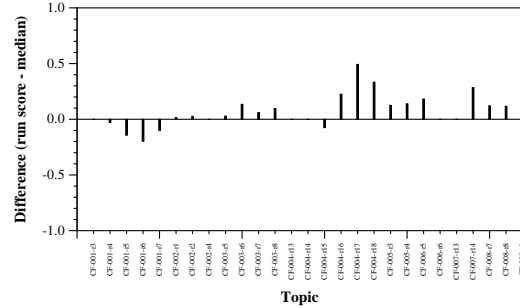
Summary Statistics	
Run ID	BM25_QAcrisis_ILP
Run type	Automatic
TREC-IS labels?	no
Streams	Twitter; Reddit; Web news
Extractive or abstractive	extractive

Manual measure means	isolated	accumulative
Comprehensiveness	0.2076	0.1543
Redundancy Ratio	0.2133	0.4527
Assessed %	0.5932	0.5932
Matched %	0.3499	0.3499
Skipped %	0.0009	0.0009
Irrelevant %	0.2026	0.2026
Unmatched %	0.4465	0.4465

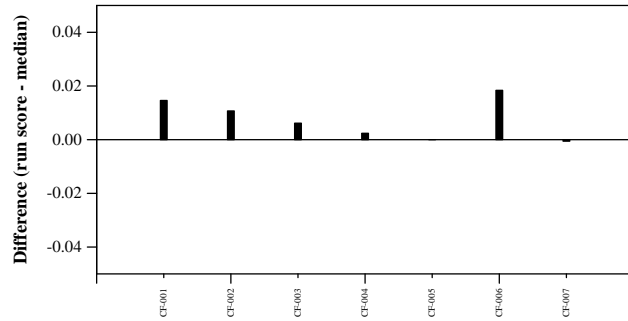
Automatic measure means	ICS	Wikipedia	NIST
ROUGE-1 F1	0.0050	0.0023	0.0100
ROUGE-2 F1	0.0010	0.0007	0.0033
ROUGE-L F1	0.0017	0.0012	0.0027
ROUGE-Lsum F1	0.0049	0.0021	0.0095
BERTscore F1	0.0098	0.0121	0.0125



Per-topic difference from median accumulative comprehensiveness



Per-topic difference from median accumulative redundancy ratio



Per-topic difference from median ROUGE-2 to ICS