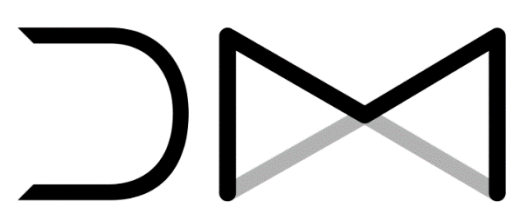


# Know Better – A Clickbait Resolving Challenge



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



DATA MANAGEMENT  
TU DARMSTADT

Benjamin Hättasch & Carsten Binnig  
LREC 2022



Dataset Statistics		
Sample Count		
Training	Dev	Test
<b>2108</b>	<b>264</b>	<b>263</b>
80 %	10 %	10%
Average Length of the Samples		
<b>Title</b>	<b>68.7</b> characters	<b>13.4</b> words
<b>Text</b>	<b>3582.2</b> characters	<b>716.6</b> words

Model using most of  
the data performs  
best across all metrics

Better than the base T5  
model but worse than  
the fine-tuned one

Baseline Ranking						
Approach	ExactMatch	Rouge-2	Rouge-L	Meteor	BLEU-2	BERTScore
Dev						
First Sentence (E)	5	7	7	6	6	8
Last Sentence (E)	5	8	8	8	8	6
Longformer Summary (S)	5	5	3	7	7	7
BART SQuADv2 (S)	4	3	4	3	3	4
S-BERT QA (E)	5	6	6	5	5	5
T5 (S)	3	4	5	4	4	3
T5 finetuned (S)	2	2	2	2	2	2
T5 augmented+finetuned (S)	1	1	1	1	1	1
Test						
First Sentence (E)	5	6	7	5	5	7
Last Sentence (E)	5	8	8	8	8	8
LongformerSummary (S)	5	5	5	7	7	6
BART SQuADv2 (S)	4	3	3	3	3	3
S-BERT QA (E)	5	7	6	6	6	5
T5 (S)	3	4	4	4	4	4
T5 finetuned (S)	2	2	2	2	2	2
T5 augmented+finetuned (S)	1	1	1	1	1	1

Extractive approaches are  
sometimes even outperformed by  
Longformer summary that does  
not utilize the teaser information

Augmentation can further  
increase the quality without  
additional manual annotation

Order of groups of  
approaches consistent  
across metrics

Shared Task: Extract the information withheld from  
the reader in a clickbait headline from the article

Baseline Scores						
Approach	ExactMatch	Rouge-2	Rouge-L	Meteor	BLEU-2	BERTScore
Dev						
Extractive Upper Bound	.0000	.1351	.2090	.1852	.0957	.1760
First Sentence (E)	.0000	.0151	.0657	.0638	.0128	.0214
Last Sentence (E)	.0000	.0070	.0386	.0455	.0058	.0553
Longformer Summary (S)	.0000	.0298	.1109	.0612	.0123	.0533
BART SQuADv2 (S)	.0038	.0565	.1030	.1164	.0508	.0476
S-BERT QA (E)	.0000	.0178	.0747	.0697	.0131	.0472
T5 (S)	.0189	.0394	.0907	.1074	.0423	.0730
T5 finetuned (S)	.0455	.0716	.1568	.1891	.0737	.1567
T5 augmented+finetuned (S)	.0720	.0870	.1846	.2296	.0910	.2089
Test						
Extractive Upper Bound	.0000	.1029	.1616	.1456	.0686	.1662
First Sentence (E)	.0000	.0187	.0517	.0582	.0125	.0514
Last Sentence (E)	.0000	.0043	.0306	.0404	.0035	.0045
LongformerSummary (S)	.0000	.0202	.0816	.0450	.0064	.0594
BART SQuADv2 (S)	.0038	.0600	.1034	.1089	.0610	.1276
S-BERT QA (E)	.0000	.0161	.0592	.0573	.0094	.0705
T5 (S)	.0114	.0275	.0849	.0952	.0250	.1124
T5 finetuned (S)	.0342	.0716	.1534	.1790	.0681	.2137
T5 augmented+finetuned (S)	.0456	.0688	.1702	.2038	.0690	.2523

Isn't this just question answering? Do we really need special models for that?			
Challenge/Pattern	Entry ID	Gold answer	Answer by fine-tuned T5
Subset (important part covered)	17669	'The Intelligent Investor' by Benjamin Graham, it advises to buy stocks when they are low and hold them and also ways to avoid huge mistakes.	The Intelligent Investor
	13522	Now, high blood pressure will be defined as 130/80	130/80
Subset (important parts missing)	10009	His Burger King food was cold	Cold
Wrong level of details	15745	They believe they'll be replaced	Fear
Repeating title phrases	21787	The author only sat on a replica inside a department store	The Iron Throne (Title was "I sat on the actual Iron Throne from 'Game of Thrones' — here's what it was like")

This  
Poster *will change*  
*your life!*

[https://link.tuda.systems/  
clickbait-resolving-challenge](https://link.tuda.systems/clickbait-resolving-challenge)

