3 Education Sector Field Study

In this section we present details of the event history field study we conducted in the educational sector setting as reported in the article.

3.1 Keywords to Optimize for & Keyword Statistics

Table W3.1 depicts the keywords (search queries) the experimental groups in the education sector field study produced optimal content for, including keyword statistics and descriptive statistics for the ranking performance of the revised machine content in the search engine. The basic keyword statistics reported in Table W3.1 include the average monthly search volume (i.e., rounded numbers of individuals that search on average per month for the keyword as provided by the search engine), the paid keyword competition, keyword length, etc. The keywords were selected such that they reflect target content and search queries in co-ordination with the educational institution running the experiments. The target keywords include both short and long tail keywords, most of them are low search volume and characterized by high keyword competition.

Table W3.1: Keywords for Field Experiments

Keyword	Descriptives								
	Avg. monthly search volume	Competition	Competition index	Keyword length	Mean revised machine ranking	Median revised machine ranking	SD revised machine ranking	IQR revised machine ranking	% of days revised machine was in ranking
quantitative marketing	10	low	-	2	6.30	5	4.80	1	100.00
quantitative marketing research	10	low	0	3	8.13	6.5	6.60	5.25	93.00
quantitative marketing program	-	-	-	3	2.03	2	0.16	0	88.40
marketing research seminar series	-	-	-	4	-	-	-	-	00.00
deep learning marketing	10	low	43	3	33.70	12	58.30	2	18.60
machine learning in marketing	10	low	16	4	84.10	98	39.10	64.5	100.00
machine learning in marketing education	-	-	-	5	1.00	1	0.00	0	34.90
digital marketing and machine learning	0	-	-	5	10.70	10	1.10	1	100.00
natural language processing in marketing	10	-	-	5	10.60	11	0.70	1	95.30
artificial intelligence in marketing	50	low	32	4	36.70	35	11.60	11.5	16.30
ai in marketing	10	low	30	3	29.00	29	0.00	0	1.00
ai in digital marketing	10	mid	36	4	77.10	78	5.70	8	67.40
ai in social media marketing	10	low	14	5	74.90	83	22.20	7.75	62.80
marketing with ai	10	-	-	3	32.50	27	22.90	14.5	46.50
marketing automation	320	mid	55	2	86.00	87	7.50	7.5	14.00
customer analytics	20	low	26	2	23.00	23	4.20	3	4.65
customer segmentation with machine learning	0	-	-	5	9.50	10	1.80	2	100.00
quantitative market research methods	10	-	-	4	8.70	10	4.20	8	86.00
business analytics education	10	-	-	3	1.00	1	0.00	0	1.00
career in marketing research	10	low	0	4	7.10	7	0.40	0	32.60
career opportunities in marketing	10	low	0	4	34.00	25	15.60	13.5	6.98
methods of marketing analytics	-	-	-	4	39.00	39	0.00	0	4.65
understanding digital marketing analytics	-	-	-	4	2.00	2	0.00	0	79.10
marketing phd career opportunities	-	-	-	4	2.00	2	0.60	0	58.10
quantitative marketing phd	10	-	-	3	4.60	1	9.00	1	93.00
doctorate PHD program in marketing	-	-	-	5	15.00	13	7.60	2.5	100.00
master program in marketing	10	mid	57	4	2.30	2	0.90	0	10.00
service marketing research	10	low	0	3	4.20	1	4.10	6	100.00
research in service marketing	-	-	-	4	4.70	5	2.30	3	100.00
academic research in service marketing	-	-	-	5	1.50	1	3.20	0	100.00
marketing institute college	-	-	-	3	10.60	3	20.30	2	100.00

Entries that display "-" mean that the search engine keyword tool did not provide specific information.

3.2 Supplemental Content Performance Tests

Table W3.2.1 reports group comparison tests for the search engine ranking performance of the experimental groups "revised machine" (printed in bold) and "human" using Wilcoxon rank sum tests. We find that the revised machine outperforms the human content generating group in terms of the number of pages that made it into the top10 search engine ranking and mean ranking performance.

Table W3.2.1: Search Engine Rankings Performance Comparison (Education Sector)

Dimension	**Group**	**Descriptives**					**Wilcoxon rank sum** [2]		
		n_p[1]	Median	(IQR)	Min	Max	W	z	*r*	*p*
Pages in ranking / day	**Revised Machine**	30	**18.00**	**(4.00)**	**12**	**22**	**908.00**	**-2.04**	**-.23**	**.041***
	Human		19.00	(2.00)	16	22				
Pages in top 10 / day	**Revised Machine**	30	**11.50**	**(3.00)**	**7.00**	**14.00**	**24.50**	**7.20**	**.82**	**.000****
	Human		5.00	(2.00)	3.00	9.00				
Mean rankings / day	**Revised Machine**	30	**17.57**	**(9.44)**	**5.60**	**30.12**	**1270.5**	**-5.79**	**-.66**	**.000****
	Human		26.22	(2.72)	19.44	30.30				

[1]n_p=number of pages per experimental group. n=77 (days); [2]Two-tailed tests; statistical significance codes: *0.05 level, **0.01 level; Compared numbers are daily aggregate numbers.

Table W3.2.2 reports quality score statistics and significance testing results for the experimental groups (i.e., revised machine versus humans), the top 10 and the lowest ranked pages on the five

quality score components. The findings are consistent with our findings from the IT service industry experiment.

Table W3.2.2: Quality Score Components Group Comparisons with Top 10 Ranked Websites (Education Sector)

Quality Score Component	Group	Descriptives				Wilcoxon rank sum[1]			
		Median (IQR)		Min	Max	W	z	*r*	*p*
Topic (s_a)	**Top 10**	**.44**	**(.20)**	**.15**	**.63**				
	Revised machine	.57	(.13)	.29	.76	205	3.72	.48	.000**
	Raw machine	.56	(.12)	.25	.76	223	3.42	.44	.000**
	Humans	.43	(.09)	.25	.64	476	-.37	-.05	.708
	Worst 10	.20	(.08)	.12	.38	803	-6.18	-.81	.000**
Keywords (s_k)	Top 10	**.47**	**(.23)**	**.09**	**.70**				
	Revised machine	.65	(.16)	.27	.85	194	3.90	.50	.000**
	Raw machine	.63	(.15)	.17	.84	190	3.97	.51	.000**
	Humans	.49	(.16)	.25	.73	393	.83	.11	.406
	Worst 10	.15	(.11)	.06	.42	805	-6.23	-.81	.000**
Uniqueness (s_d)	Top 10	**.95**	**(.06)**	**.78**	**.99**				
	Revised Machine	.94	(.08)	.73	1.00	527	-1.13	-.15	.258
	Raw machine	.90	(.09)	.72	.99	623	-2.57	-.33	.010*
	Humans	.99	(.05)	.07	1.00	234	3.19	.41	.001**
	Worst 10	.94	(.03)	.86	.98	459	.35	.05	.724
Readability similarity (s_r)	Top 10	**.57**	**(.07)**	**.46**	**1.00**				
	Revised Machine	.73	(.25)	.47	1.00	135.5	4.64	.60	.000**
	Raw machine	.78	(.23)	.45	1.00	85	5.39	.70	.000**
	Humans	.47	(.54)	.00	1.00	544.5	-1.39	-.18	.165
	Worst 10	.41	(.23)	.15	.70	726	-4.41	-.57	.000**
Naturality similarity (s_n)	Top 10	**.57**	**(.08)**	**.44**	**.75**				
	Revised Machine	.54	(.50)	.17	1.00	461	-0.15	-.02	.876
	Raw machine	.67	(.50)	.33	1.00	393	.84	.11	.403
	Humans	.17	(.25)	.00	1.00	780	-4.89	-.63	.000**
	Worst 10	.41	(.16)	.20	.88	725.5	-4.40	-.57	.000**

[1] Two-tailed tests; statistical significance codes: *0.05 level, **0.01 level;

Appendix References

Baayen RH, Shafaei-Bajestan E (2019) Analyzing linguistic data: A practical introduction to statistics. Package 'languageR'. Version 1.5.0. *CRAN*. Accessed May 20, 2019, https://cran.r-project.org/web/packages/languageR/languageR.pdf

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A, (2018) "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*. 3(30). https://doi.org/10.21105/joss.00774

Berger J, Sherman G, Ungar L (2020b) TextAnalyzer. Accessed November 11, 2020, http://textanalyzer.org

Bronnenberg BJ, Kim JB, Mela CF (2016) Zooming in on choice: How do consumers search for cameras online? *Marketing Science*. 35(5):693-712.

Danaher PJ, Mullarkey GW, Essegaier S (2006) Factors affecting website visit duration: A cross-domain analysis. *Journal of Marketing Research*. 43(2):182-194.

Edelman B, Zhenyu L (2016) Design of search engine services: Channel interdependence in search engine results. *Journal of Marketing Research*. 53(6):881-900.

Flanigan, AJ, Metzger, MJ (2007) The role of site features, user attribtues, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*. 9(2):319-342. https://doi.org/10.1177/1461444807075015

Jerath K, Ma L, Park YH (2014) Consumer click behavior at a search engine: The role of keyword popularity. *Journal of Marketing Research*. 51(4):480-486.

Kamoen N, Holleman B, Bergh H (2013) Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability. *Survey Research Methods*. 7(3):181-189.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science*. 37(6):930-952.

Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao ELT, Palma MA (2020) Basic robust statistics. Package 'robustbase'. Version 0.93-6. *CRAN*. Accessed May 20, 2020, https://cran.r-project.org/web/packages/robustbase/robustbase.pdf

Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) Linguistic inquiry and word count: LIWC2015. Austin, TX: Pennebaker Conglomerates. Accessed November 1, 2020, www.LIWC.net.

Pitler E, Nenkova A (2008) Revisiting Readability: A unified framework for predicting text quality. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 186-195.

Radford A, Narasimhan K, Salimans T, Sutskever I (2018) Improving language understanding by generative pre-training. OpenAI.

Roberts C (2010) Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*. 54(1):43-56.

Rocklage MD, Rucker DD, Nordgren LF (2018) Persuasion, emotion and language: the intent to persuade transforms language via emotionality. *Psychological Science*. 29(5):749-760.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomze AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*. 1-15.