2.3.3 Content Ranking for Sub-Keywords Assessment

In SEO, content is usually optimized for a single main keyword (search query). However, in search engine advertising (SEA) ads and bids are often optimized for multiple keywords at once. Thus, a company could potentially benefit from optimizing SEO content for multiple keywords simultaneously. That is why we conducted an additional experiment for our IT service industry study to explore how well the experimental groups' SEO content ranks for related keywords. We identified the latter by analyzing the word distributions of the top 10 search engine ranked content for the 19 main keywords specified in Table W2.1 and extracted the most frequent keywords and groups of words, yielding 207 related keywords. For example, when analyzing the top 10 search results for the keyword "IT assessment", we find the (most frequently occurring) following related keywords based on their word distributions: "IT assessments", "business continuity", disaster recovery", "security assessment", "assessment services", "IT assessment services", "risk security assessment", "information technology assessment", "disaster recovery plan". After scraping the search engine rankings for these 207 related keywords, we find that the revised content machine ranks substantially better and more often for related keywords than the competing human groups (Table W2.3.3). For example, the revised machine ranked for 34 related keywords and occurred in the top 10 results six times. The median search engine ranking of the revised machine is 23.

Based on this auxiliary study, we conclude that the method seems to perform surprisingly well for related keywords as well. In contrast to relying on heuristics such as keyword density, the fine-tuning process of our semi-automated algorithm appears to not only capture the overall topic but also related sub-topics within the content. Thus, in the process of

generating content for a specific keyword, our content also performs reasonably well in terms of search engine rankings for topic-related sub-keywords for which it was not primarily optimized.

Table W2.3.3: Ranking Performance of Content for Related Keywords

	Descriptives			
Group	Median (IQR) search engine ranking		Total number of ranked pages	Number of pages ranked in top10
Revised machine	**23**	**(23.00)**	**34**	**6**
Real SEO Experts	26	(35.50)	4	1
Quasi Experts	188.5	(78.25)	4	0
Novices	68	(27.50)	3	0

Desriptives for achieved rankings per experimental group for topic-related sub-keywords extracted from the top 10 ranked pages (207 sub-keywords, and 51,995 total ranked pages).

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.