The Emotion Profile of Web Search

Gabriella Kazai
gkazai@microsoft.com
Microsoft
UK

Paul Thomas
pathom@microsoft.com
Microsoft
Australia

Nick Craswell
nickcr@microsoft.com
Microsoft
US

ABSTRACT

Emotions are an essential part of most human activities, including decision-making. Emotions arise in response to information, e.g., presented in web pages, and are also expressed in the words used to convey that information in the first place. In this paper, we study the emotion profile of retrieved and clicked web search results towards the goal of better understanding the role of emotions in web search. Using click logs from a four-month period, up to the end of January 2019, we examine the emotions associated with search results and contrast them to the emotions of clicked results, taking rank and relevance into account. Emotions are assigned to web pages based on two lexicons: SentiWordNet (positive, negative and objective sentiments) and EmoLexData (afraid, amused, angry, annoyed, don't care, happy, inspired, and sad emotions). We look at the sentiment/emotion profiles of search results grouped around a set of controversial and mundane topics and hypothesise that users are more likely to click emotionally charged results than emotionless results, both in general, and in particular when their query relates to controversial topics.

CCS CONCEPTS

• **Information systems → Sentiment analysis**.

KEYWORDS

Emotions in web search; emotional profile; click behaviour

ACM Reference Format:
Gabriella Kazai, Paul Thomas, and Nick Craswell. 2019. The Emotion Profile of Web Search. In *42nd Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331314

1 INTRODUCTION

Following the seminal work of Kahneman and Tversky [6], many psychologists argue that humans are often less driven by rational thinking, logical reasoning and the consideration of statistics or objective facts, and—as inherently emotional creatures—are more likely to think and act by relying on emotions, at least unconsciously [12]. Indeed, emotions are known to play an integral part of information search processes as they affect a searcher's attention, memory, performance, and judgments [8]. As a result, emotion has been the subject of numerous studies, investigating the role of users' emotions in their willingness to search, their search strategies and search performance, as well as studying the affect of the search process on the users' emotions.

However, while considerable research has focused on the study of searchers' emotions, little attention has been devoted to examining the emotional makeup of the web pages themselves that are served to users by search engines or the emotion profile of the subset of web pages that users then select for further examination. What can be said in general about the emotional charge of search engine result pages (SERPs) returned to users? For example, can we observe a general trend whereby more emotional results are ranked at the top of SERPs? What about the relationship between users' click behaviour and the emotional charge of clicked documents: can we see evidence of users favouring more emotional results?

In this paper we take a first step to studying the emotion profile of web search towards better understanding the role of emotions in search and ultimately in human decisions informed through the search. For our study, we make use of the query logs of the Bing search engine; a sentiment and an emotion lexicon, and conduct a preliminary analysis of the search results' sentiment/emotion signatures across a number of segments: clicked vs not-clicked results, results returned for controversial vs mundane topics, relevant vs irrelevant results, etc. Our findings show that, in our sample data, there is a difference in the emotion profile of search results returned for mundane and controversial topics, the latter exhibiting more emotions such as anger and annoyance. We also see that while results at the top of the SERP use less emotive language, searchers are drawn to click on results preferentially in ways that correlate with expressed sentiment and emotion.

2 RELATED WORK

Emotions express what stimuli (including information) mean to us and they influence our reaction to the stimuli [4]. As such, emotions provide a primary response mechanism that can focus our attention and impact our perception, thinking and decisions [8].

The sentiment, and more recently, the emotional content of online texts such as news articles, blogs, social media posts or user reviews have been of great interest for a range of applications including opinion mining for products, entities, or events, e.g., elections. Typically, sentiment analysis methods aim to derive the sentiment/emotion of previously unseen texts by exploiting the use of words or texts with explicit affective labels [7]. Lexical resources, such as SentiWordNet [1], organise labelled affective words into taxonomies, enabling for sentiment analysis to be performed directly via lookups or indirectly by training classifiers [3]. Recently, approaches employing distributional representation and Deep Learning for emotion detection in texts have also been proposed, see Zhang et al. [14] for a review. In our current work we

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07...$15.00
https://doi.org/10.1145/3331184.3331314

opt to use two open-source lexicons as a way of attaching sentiments/emotions to web pages to enable our analysis.

In the context of affective aspects in information seeking behaviour, since the early research of Kuhlthau et al. in the 1990s, numerous user studies have shown that emotions affect users' willingness to search, their search strategies and search performance, and that users' emotions are in turn affected by the search process, the system's performance, the user's interest in the process and documents, and other variables [8].

However, as far as we are aware, very little attention has been focused on the emotional content of search results in web search and its influence on users' click behaviour. Lopatovska and Mokros [9] measured the quality and intensity of users' feelings by collecting "willingness to pay" and "experienced utility" ratings for a handful of retrieved documents. Greving and Sassenberg investigated how threat influences web search and found that when under threat, users' attention is directed more to positive pages and more positive information is acquired and retained [5]. Closest to our work is that of Demartini and Siersdorfer [3], who extracted sentiments from web pages and compared up to 50 results from 3 commercial search engines on 14 controversial queries. They saw no significant differences between the search engines, but noted differences in the sentiments expressed in retrieved pages for different queries. They also found a relation between rank and sentiment: results ranked on top were, on average, more positive.

3 METHODS AND DATA

We study emotions expressed by the titles and snippets of search results by analyzing Bing's query logs from a four months period.

Data. We start by selecting a set of mundane and controversial topics, and a handful of positive ('sanity-check') topics that we expect to be associated with positive emotional content: see Table 1. The positive topics are only used for checking that the lexicon approach is reasonable. The controversial topics are those used by Demartini and Siersdorfer [3], expanded with more recent topics. The mundane topics reflect less controversial, everyday or scholarly concepts, expecting that these would elicit less expressed emotions by web page authors. We deliberately pick topics based on intuition and prior to any analysis to avoid a circular definition.

We extract user queries, SERPs and user clicks (up to rank 5; results further down elicit much fewer clicks [2]) from a four-month period, between 1st of October 2018 and 31st of January 2019, from Bing's query logs. We keep only queries that contain our topic words. As a result, we obtain over 16 billion query-URL impressions (88% of which relate to mundane, 9% to controversial and 3% to positive topics) with over 1 billion clicked URLs.

For each query-URL pair, we also obtain a document representation, consisting of the title and the snippet that was returned to the user on the SERP. Finally, we merge relevance labels for query-URL pairs when available: converting the original judgments to binary relevant or irrelevant labels.

Assigning emotion labels. To assign sentiments/emotions to the search results, we follow a lexicon-based approach and make use of the following two open-source lexicons:

- *SentiWordNet* [1] is a lexicon in which a triple of sentiment values, positive (pos), negative (neg), and objective (obj), is assigned to each set of synonymous words in WordNet. Each value is in $[0, 1]$ and the three values sum to 1. For example, the adjective "perfect" has a score of pos=0.625, neg=0.125, and obj=0.25. The lexicon contains about 200k words.
- *EmoLexData* is a lexicon built by Song et al. [11] that contains about 32k words, each annotated with a vector of eight emotion values: afraid, amused, angry, annoyed, dont_care, happy, inspired and sad. The lexicon was built based on 31k news articles from rappler.com, where users visiting the site provided the original emotion labels.

While such dictionary based approaches can suffer from low recall, they usually attain high precision. We opt for this approach, rather than alternative methods that involve training a classifier, for its simplicity and ease of interpretation. In addition, as both lexicons are much more extensive than other often-used lexicons, e.g., LIWC or WordNet-Affect[1], we expected recall to be less of an issue. Indeed, each of the lexicons on average matched 93–94% of terms in our document representations. Unmatched words were associated with the sentiment/emotion values of (pos=0, neg=0, obj=1)/(dont_care=1 and 0 otherwise), respectively.

Given sentiment/emotion vectors for individual words, the sentiment/emotion of a search result is derived as the element-wise average of its word vectors. The sentiment/emotion of words with multiple part-of-speech roles are also averaged.

4 ANALYSIS AND RESULTS

We analyze the four months of log data for our selected topics by comparing the emotion profile of the search results' titles and snippets, segmented by rank position, relevance, clicks and topic.

Overall Emotion Profile. First we derive an overall emotion profile by averaging the emotion vectors associated with all documents across all observed SERPs in our log data. We can average two ways: 1) micro-averaging across all queries and SERPs would result in an emotion profile that represents the "average experience" but is dominated by the most popular, so-called head queries or 2) we can first average emotions per topic and then macro-average across topics. Rows 1 and 2 in Table 2 show the obtained overall emotion profiles. The sentiment scores suggest that most of the search results are objective or emotionless. However, based on the finer-granularity emotion scores, we see that the happy emotion dominates, followed by dont_care and inspired. The difference between the micro and macro averages shows the impact of head queries, such as "facebook" or "craigslist", which appear to be more neutral (higher dont_care emotion contribution). For the rest of the analysis, we report only macro-averages and only over the mundane and controversial topics, except for row 12.

Emotions across Rank Positions. To examine the relationship between emotion and rank positions, we segment our data by rank and average the emotion vectors at each rank: see rows 3–7 in Table 2. Both the sentiment and emotion vectors show that search results at the top of the ranking are less emotional than those further down

[1] http://liwc.wpengine.com/; http://wndomains.fbk.eu/wnaffect.html

Table 1: Topics in our "mundane", "controversial", and "positive" sets

Mundane	Controversial	Positive
bank, bookcase, calculus, cognitive, craigslist, definition, equilibrium, facebook, football, google, jobs, maths, microsoft, nfl, school, shoes, thermodynamics, weather, windows	abortion, animal test, anorexia/eating disorder, economy, employment, gay/lesbian, genetic clon, gun control, hitler, immigrat, islam, marijuana, nazi, obama, terrori, trump, vegetari, war	christmas/santa, cute cat, marriage/wedding

Table 2: Emotion Profiles: sentiment/emotion vectors sum to 100%; statistical significance (two tail t test) $^{\dagger i}p < 0.05$,$^{\ddagger i}$ $p < 0.01$ vs row i; with the exception of rows 1, 2, and 12, all statistics are calculated over the mundane and controversial topics only

		Sentiment			Emotion							
		pos	neg	obj	afraid	amused	angry	annoyed	dont_care	happy	inspired	sad
1	Micro-avg all	3.23	2.30	94.46	3.36	6.23	6.02	4.74	30.83	34.04	7.51	7.26
2	Macro-avg all	4.40	3.31	92.29	4.82	7.92	7.64	6.86	19.92	36.18	10.18	6.47
3	Rank 1	3.78	2.97	93.26	4.82	7.84	7.48	6.45	24.54	32.74	9.75	6.38
4	Rank 2	$4.37^{\dagger 3}$	3.33	$92.30^{\dagger 3}$	5.11	7.76	7.68	7.09	$18.54^{\ddagger 3}$	$36.63^{\ddagger 3}$	10.61	6.58
5	Rank 3	4.79	3.64	91.57	5.05	7.43	8.16	7.35	$14.79^{\ddagger 4}$	$39.77^{\ddagger 4}$	10.72	6.73
6	Rank 4	4.80	3.66	91.54	5.10	7.49	8.13	7.30	14.66	39.81	10.82	6.68
7	Rank 5	4.83	3.68	91.48	5.11	7.49	8.14	7.29	14.36	39.86	11.00	6.76
8	Not clicked	4.24	3.28	92.48	4.92	7.59	7.74	6.82	20.38	35.77	10.25	6.53
9	Clicked	$4.78^{\dagger 8}$	3.50	91.72	4.95	7.31	8.44	7.15	$14.94^{\ddagger 8}$	$39.54^{\ddagger 8}$	10.87	6.82
10	Mundane	4.08	3.06	92.86	4.21	7.51	6.23	6.25	22.56	36.77	9.63	6.84
11	Controversial	$4.48^{\dagger 10}$	$3.53^{\dagger 10}$	$91.99^{\dagger 10}$	$5.66^{\dagger 10}$	7.63	$9.42^{\ddagger 10}$	$7.45^{\dagger 10}$	$17.40^{\ddagger 10}$	35.21	10.97	6.25
12	Positive topics	$5.40^{\ddagger 10}$	3.46	$91.14^{\dagger 10}$	4.06	$10.62^{\dagger 10}$	6.52	7.07	18.97	37.53	9.36	5.86
13	Irrelevant	4.26	3.23	92.51	4.38	7.42	8.93	7.75	16.44	39.39	8.64	7.04
14	Relevant	4.87	3.67	91.46	5.09	7.05	8.03	6.89	16.48	38.74	$11.40^{\dagger 13}$	6.33

the ranking. We observe significant correlations with rank for positive, negative, objective sentiments, annoyed, dont_care, and happy emotions ($\rho = 0.38, 0.28, -0.37, 0.20, -0.62, 0.44$; all $p \ll 0.01$), and inspired ($\rho = 0.16$; $p < 0.05$). We believe this is a side-effect of ranking decisions such as preferring official, authoritative sources. Indeed, the reduced emotional charge of the top result, compared with lower ranks, is even stronger for head queries, where top results tend to be emotionless navigational results. When limiting to controversial topics, we observe no statistical differences between ranks 1 and 2 (but still do between ranks 2 and 3). The effect seen here is in contrast to the findings of Demartini and Siersdorfer [3] where results ranked on top, on average, were more positive. While our dataset also includes mundane topics, the trend of less emotional results at top ranks still holds for our controversial topics (significant correlations with rank for positive and objective sentiments ($\rho = 0.31, -0.32$; $p < 0.01$), dont_care and happy emotions ($\rho = -0.63, 0.33$; $p \ll 0.01$), and amused ($\rho = -0.21$; $p < 0.05$).

Clicked vs Not-Clicked. Rows 8–9 in Table 2 contrast the emotion profiles of clicked vs not-clicked search results. We can see a clear user preference for emotional results: clicked results are significantly more positive and happy than not-clicked results. This suggests that despite the search engine ranking emotionless results higher up the ranking, users are more likely to click on emotional results, especially those expressing happiness.

Controversial vs Mundane Topics. Grouping by topic type, rows 10–12 of Table 2, confirms our expectations that controversial topics surface search results that are more emotionally charged than mundane topics. We find statistically significant differences in positive, negative and objective sentiments, in afraid and annoyed ($p < 0.05$), angry and dont_care emotions ($p < 0.01$) for controversial topics compared with mundane topics. As expected, the positive topics' emotion profile shows high levels of positive sentiment, but, interestingly, it is amused, not happy, that shows a significant increase compared to mundane topics.

Relevant vs Irrelevant. So far we saw that although SERPs tend to contain less emotional results at the top, users are more likely to click emotionally charged results. Next, we investigate if relevant and irrelevant results differ based on their emotion profile: if relevant results happen to be more emotional, this could explain—at least some of—the observed click preferences. This analysis is based on the subset of 5 million query-URL pairs with a relevance label in our dataset. Rows 13–14 in Table 2 show no significant sentiment differences between relevant and irrelevant results, with the exception of relevant results expressing higher inspired emotions than irrelevant results.

Propensity to Click. Next, we ask whether particular expressions of sentiment or emotion encourage searchers to click on a snippet: that is, whether the probability of a click correlates with our sentiment or emotion variables. To examine this, we take a 2% sample of our result-level data, stratified by topic, rank, relevance, and whether the result was clicked. (That is, for each combination of these four variables, we take a 2% uniform sample.) This gives

Table 3: Contributions of rank, relevance, and sentiment to observed clicks. Effects in log odds scale, all effects $p \ll 0.01$.

	Variable	Effect	
1	Rank	−1.13	
2	Relevant	1.44	
3	Irrelevant	0.00	(baseline)
4	Pos	6.64	
5	Neg	−5.43	
6	Relevant + pos	19.48	
7	Relevant + neg	−11.56	
8	Irrelevant + pos	0.00	(baseline)
9	Irrelevant + neg	0.00	(baseline)

329 million data points. We then model whether each result was clicked using logistic regression.

Table 3 summarises the best-fit model incorporating rank, relevance, sentiment, and the interactions between relevance and sentiment[2]. The effects are given on a log-odds (logistic) scale, so effects > 0 correspond to an increased probability of a click and effects < 0 to a decreased probability. All effects are statistically significant ($p \ll .001$).

As previously observed [2], click rate decreases with rank due to user biases, search engine effectiveness, and searchers reading from top down (row 1). We also see, as expected, more clicks on relevant documents than on those judged irrelevant (rows 2–3).

Rows 4–5 show the effect of sentiment on the probability that a result will be clicked. A purely positive document title and snippet would increase the (log) odds of a click by 6.64 points, even after accounting for rank and relevance, while purely negative text would reduce it by almost as much. It appears that searchers are drawn to positive sentiment on a SERP.

We also see interactions between relevance and sentiment. Of documents which are topically relevant, those which appear positive are much more likely to be clicked (19.48 points on top of the 6.64 from the main effect) than those which appear negative (−11.56 atop −5.43). In other words, for relevant results the effect of sentiment is greatly exaggerated compared to irrelevant results.

A separate model (not shown here), with terms for rank, relevance, and emotion, demonstrated similar effects. Higher-ranked and relevant results are more likely to be clicked (effect of rank −1.16, relevance 1.87), as are results which seem happy (0.21), inspired (0.50), or angry (1.17). Results get fewer clicks if they seem neutral (dont_care, −1.91), afraid (−1.03), or amused (−0.87).

5 CONCLUSIONS

This paper examines the emotion profile of web search for a variety of mundane and controversial topics. For those topics, we analyzed approximately 16 billion search impressions, assigning each of the top-5 search results for each impression an emotional profile using SentiWordNet, and a separate emotional profile using EmoLexData.

We identify a number of patterns in emotion profiles across different segments of our data: Ranking criteria seem to have a side-effect of less emotional content near the top of the ranking, on average. Surprisingly, the clicked results show the opposite pattern, with clicked results having significantly higher positive sentiment and happiness compared to not-clicked results. We also see that controversial topics lead to statistically significantly more afraid, angry, annoyed emotional results than mundane topics. Finally, we find that the profile of relevant documents shows significantly higher levels of inspired emotions than irrelevant documents.

We used regression to understand how click decisions are influenced by rank, relevance and sentiment. After controlling for rank and relevance, we find that positive results are more likely to be clicked, and this is particularly true for results that were also judged relevant. This preference for emotionally-positive results could be loosely related to White's finding that positive (confirmatory) responses were favored for yes-no medical questions [13], although in our case we are not limited to yes-no questions and our notion of positive is based on sentiment. The preference for the positive could also be related to work by Greving and Sassenberg [5], who found that in a negative affective or motivational state, attention is automatically allocated to positive information.

In future we could incorporate more lexicons or go beyond lexicon-based approaches for emotion analysis, expand our initial set of topics and consider general web search traffic, incorporate the emotion profile of the web page itself, and explicitly consider user intentions, e.g., users looking to be entertained [10]. This could confirm patterns of behavior around emotion, and lead to the harder questions of why users behave in these ways and how search systems should be designed to better serve user needs, given these preferences or biases. For example, could or should search engines help to make users aware of emotional biases?

[2] "Objective" of course strongly anticorrelates with the sum of "positive" and "negative", so we drop this from the model.

REFERENCES

[1] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. Int. Conf. Language Resources and Evaluation.*

[2] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proc. WSDM.* 87–94.

[3] Gianluca Demartini and Stefan Siersdorfer. 2010. Dear search engine: What's your opinion about...?: Sentiment analysis for semantic enrichment of web search results. In *Proc. Int. Semantic Search Workshop.* Article 4, 7 pages.

[4] Nico H Frijda. 2007. *The laws of emotion* (2nd. ed.). Routledge, New York, NY.

[5] Hannah Greving and Kai Sassenberg. 2015. Counter-regulation online. *Comput. Hum. Behav.* 50, C (Sept. 2015), 291–298.

[6] Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (March 1979), 263–291.

[7] Evgeny Kim and Roman Klinger. 2018. A survey on sentiment and emotion analysis for computational literary studies. *CoRR* abs/1808.03137 (2018). arXiv:1808.03137

[8] Irene Lopatovska. 2014. Toward a model of emotions and mood in the online information search process. *JASIST* 65, 9 (2014), 1775–1793.

[9] Irene Lopatovska and Hartmut B. Mokros. 2008. Willingness to pay and experienced utility as measures of affective value of information objects: Users' accounts. *Inf. Process. Manage.* 44, 1 (2008), 92–104.

[10] Yashar Moshfeghi and Joemon M. Jose. 2013. On Cognition, Emotion, and Interaction Aspects of Search Tasks with Different Search Intentions. In *Proc. WWW.* 931–942.

[11] Kaisong Song, Wei Gao, Ling Chen, Shi Feng, Daling Wang, and Chengqi Zhang. 2016. Build emotion lexicon from the mood of crowd via topic-assisted joint non-negative matrix factorization. In *Proc. SIGIR.* 773–776.

[12] Kirsten G. Volz and Ralph Hertwig. 2016. Emotions and decisions beyond conceptual vagueness and the rationality muddle. *Perspectives on Psychological Science* 11, 1 (Jan. 2016), 101–116.

[13] Ryen W White. 2013. Beliefs and biases in web search. In *Proc. SIGIR.* 3–12.

[14] Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 8, 4 (2018).