On User Interactions with Query Auto-Completion

Bhaskar Mitra, Milad Shokouhi, Filip Radlinski, Katja Hofmann
Microsoft
Cambridge, UK
{bmitra, milads, filiprad, katja.hofmann}@microsoft.com

ABSTRACT

Query Auto-Completion (QAC) is a popular feature of web search engines that aims to assist users to formulate queries faster and avoid spelling mistakes by presenting them with possible completions as soon as they start typing. However, despite the wide adoption of auto-completion in search systems, there is little published on how users interact with such services.

In this paper, we present the first large-scale study of user interactions with auto-completion based on query logs of Bing, a commercial search engine. Our results confirm that lower-ranked auto-completion suggestions receive substantially lower engagement than those ranked higher. We also observe that users are most likely to engage with auto-completion after typing about half of the query, and in particular at word boundaries. Interestingly, we also noticed that the likelihood of using auto-completion varies with the distance of query characters on the keyboard.

Overall, we believe that the results reported in our study provide valuable insights for understanding user engagement with auto-completion, and are likely to inform the design of more effective QAC systems.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval

Keywords: Query suggestions, auto completion

1. INTRODUCTION

Most modern search engines implement some form of Query Auto-Completion (QAC), where users are shown possible completions for their query from the moment they type the first character. This sort of system is assumed to make it easier for users to enter effective search queries, in particular by avoiding common mistakes such as misspellings and using particularly ambiguous queries.

However, the details of user's interactions with QAC systems are poorly understood. In this paper we study some of the latent factors besides query suggestion quality that influence user engagement with QAC systems. Specifically, we analyze the logs from Bing, a large commercial search engine, to identify conditions under which users are more likely to engage with a QAC system.

Query logs capture the interactions of users with search systems and provide valuable insights about how users search. Therefore, perhaps it is not surprising that query logs have been studied extensively to model user interactions with different components of search engines, such as search results [8], vertical results [1], ads [4] and related searches [2]. In this paper, we present the first large-scale query-log analysis on QAC.

In the first part of our analysis, we investigate how QAC usage varies with suggestion position and the topical class of a user's intent. In the second part, we focus on differences in QAC usage with respect to various attributes of the partial query already typed – typically referred to as *query prefix* – at the point of engagement .

SIGIR '14 , July 06 - 11 2014, Gold Coast , QLD, Australia

http://dx.doi.org/10.1145/2600428.2609508

2. RELATED WORK

QAC systems produce a list of suggested queries given a prefix usually using exact-prefix matching. The ranking order of suggestions is typically static, and based on past query frequency [3, 5]. However, recent work has also considered ordering suggestions by predicted future query popularity [12] or personalizing QAC to the specific users [3, 11, 13]. Beyond the standard desktop interface, Kamvar and Baluja [9] reported that auto-completion on mobile devices could save users from typing up to 46% of their query characters.

However, there have been few studies on understanding user interaction with QAC. Seeking to model user behavior, Kharitonov et al. [10] proposed a model of users as interacting with QAC suggestions from top to bottom, looking for a specific suggestion they have in mind. However, this does not address the question of *when* during query formulation the users will actually choose to do this. In terms of ranking quality, the only previous work that directly measures the effect of changes to QAC was a user study involving an enterprise search system [7]. The authors find that improving auto-completion by adding faceted-navigation features significantly reduces search time when users look for known items. Again, the question of when precisely this interaction happens is not addressed.

3. ANALYSIS

We sampled approximately 1.6 million queries from Bing's search query log for the first week of February 2014. To model the user goal, we follow previous work on QAC [3, 11] and assume that the submitted query is the target query the user wanted to formulate. We note that while the actual QAC suggestions likely influence the submitted query, in our analysis we did not consider the degree to which ranking quality affected the user's choice to interact with the QAC system. Also, the only form of engagement with QAC that we consider involves the user submitting a query proposed by the QAC

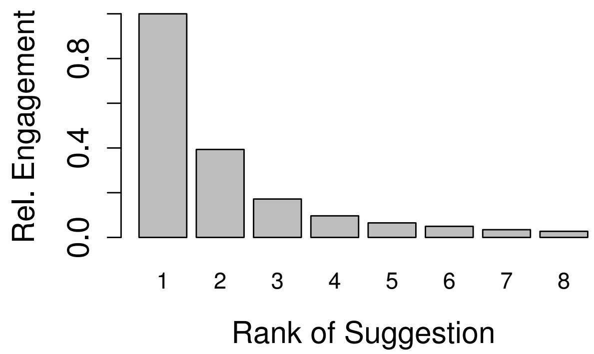

Figure 1: The distribution of QAC engagement by rank of the suggestions. Due to proprietary nature of the dataset, the numbers are normalized with respect to the first position and others rescaled proportionally.

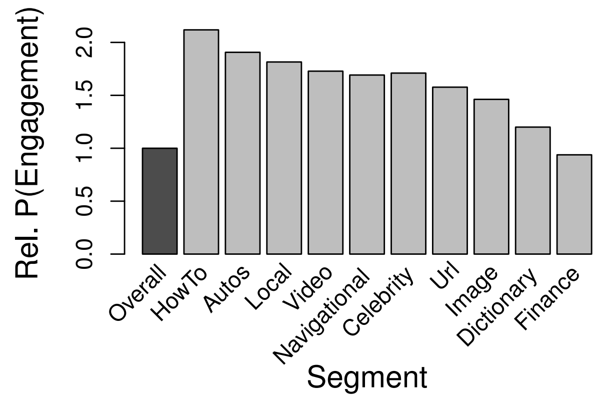

Figure 2: The relationship between query class (of the final issued query) and probability of engagement with QAC. We see that a user's is likelihood of engaging with QAC varies substantially with query class. As before, we only present the ratio of the probability of QAC engagement per segment to the average overall probability.

system. In particular, this ignores other forms of potential QAC usage – such as users simply observing the suggestions and copying the correct spelling of query terms during query formulation.

We start by investigating the overall usage patterns of QAC based on suggestion positions and types of intent. We then focus on usage differences with respect to specific attributes of the query prefix.

3.1 QAC usage by type and rank

Suggestion rank. Figure 1 shows the distribution of QAC engagement by suggestion rank. Given the proprietary nature of the Bing's dataset, we do not report the absolute probabilities of QAC engagement in this paper and instead show the ratio between the probability of QAC engagement for each position and the first position.

We see a clear decrease in engagement with lower ranks of QAC suggestions. This dramatic rank effect (with the second position receiving less than 40% of the engagement of the top suggestion) may be caused by the lower-ranked suggestions being less relevant, or by strong position bias, with users assuming that top-ranked suggestions are better. Previous work has found that in the case of Web search users often select the top result when a target result is hard to find [6]. The relative contribution of these two factors in QAC remains to be determined.

Intent type. Does QAC engagement vary based on the type or topical category of the query? To answer this question we ran a set of available in-house trained query classifiers on the submitted queries and computed the probability of QAC engagement for each segment corresponding to each classifier. As before, we report only the relative ratio (and not the absolute values) of the probability of QAC engagement for each segment to the overall probability of QAC engagement across all queries.

Our data shows that QAC engagement varies with query class. For instance, there is a higher probability of QAC engagement for How-To, Local and Navigational queries, and lesser than average for query segments such as finance. We hypothesize that for popular queries, such as navigational and famous celebrities, the search engine is more likely to return relevant suggestions. This may explain the higher than average engagement with QAC in these segments.

3.2 QAC usage by prefix attributes

Word boundaries. Given that a user engages with QAC, when will this occur? We now ask if there is more or less user engagement with QAC at different parts of the query. Figure 3 shows that within a word the highest user engagement with QAC suggestions happens after the user has typed the first 3 characters, and that it decreases with further increase in the number of characters typed for the current word. While the relative frequency of words of different lengths in the search logs may partially explain this pattern, we observe the same pattern of decreasing engagement with decreasing distance from the end of the word in Figure 4 in the range where 4 to 1 characters in the word remain to be typed. We hypothesize that this indicates that the closer the user is to typing the full word, the less likely he/she is to engage with QAC for completing the word. Interestingly, we also note the apparent *bursty* nature of typing at the very beginning of each word, where users are less likely to engage with QAC after typing just the first or the second characters of a word. We hypothesize that once users start typing a word, they often type the first three characters in a semi-automatic fashion.

The user preferences also become obvious when we look at the engagement with respect to distance from the end of a word. As shown in Figure 4, about 38% of QAC engagements happen after the user has typed the last character of a word and an additional almost 15% of engagement happens after the user typed the space character immediately following a word. In other words, at this point they are most likely to select the completion of their query in terms of just selecting additional words to be appended to what they have typed so far.

Fraction of query typed. What fraction of the query do users generally type before engaging with QAC suggestions? Based on our analysis, QAC engagement most frequently happens after the user has typed slightly more than 50% of the query, as shown in

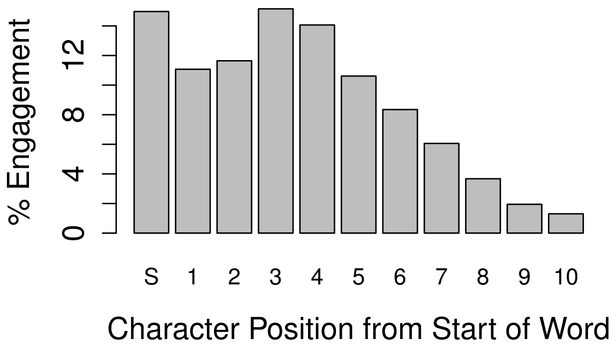

Figure 3: Percentage of engagement with QAC based on the number of characters typed in the current word. "S" denotes the character that separates the current word from the previous.

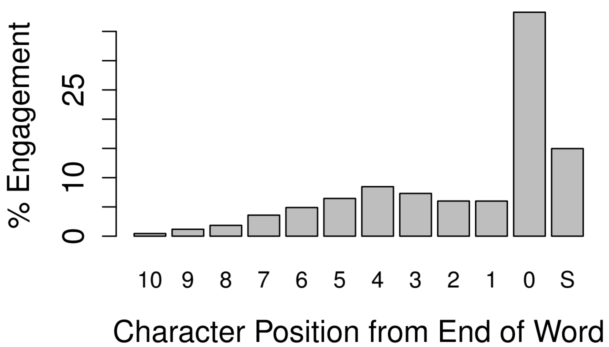

Figure 4: Percentage of engagement with QAC based on the number of characters that have not yet been typed in the current word. "S" denotes the character that separates the current word from the next.

Figure 5. Interestingly there is also a substantial number of instances where the user typed the full query but decided to use QAC for submission instead of pressing the return key or clicking the search button.

Keyboard distance. Does the distance the finger needs to travel for the next keystroke influence QAC engagement? To test this hypothesis, we considered all pairs of keystrokes where the first key corresponds to the last character of the prefix already typed by the user at the point of engagement with QAC and the second key corresponds to what would have been the next character based on the submitted query. We restricted our analysis to only pairs of characters where both were in the English alphabet for simplicity. We then computed a keyboard distance between any two given keys based on the minimum number of adjacent keys that the finger must travel over to reach one from the other assuming a QWERTY keyboard layout. For example the keyboard distance for the {a,s} pair is 1 and {a,c} is 3 (based on the path a→s→d→c) by this definition.

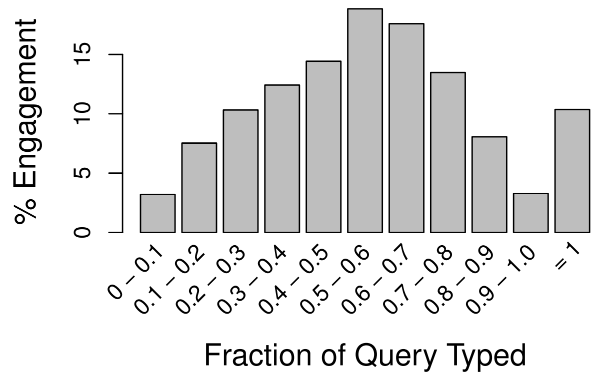

Figure 5: Fraction of the query that is typed by the user before engaging with QAC.

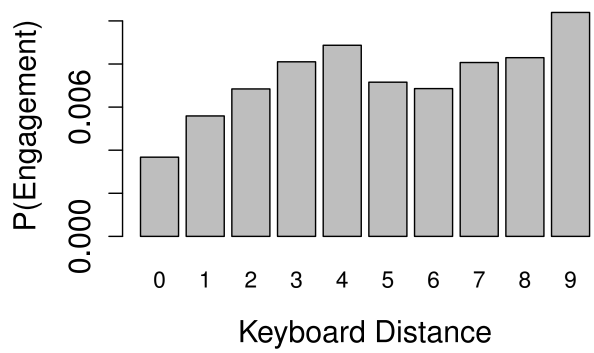

Figure 6: Effect on keyboard distance between the last character typed and the next character that the user would have had to type had they not engaged with QAC on user engagement.

Figure 6 shows a monotonic increase in the probability of engagement with keyboard distance up to a distance of 4 keys. The 4 keys distance is interesting as it is the maximum distance a user's finger must travel while typing on a QWERTY keyboard[1] assuming the use of two hands and that each hand covers half the layout area of the keyboard. This means that even though for the pair {a,p} the keyboard distance according to our definition is 9, the actual distance any individual finger of the user needs to travel will be less than 9 as the user will most likely employ fingers on different hands to reach the two keys. We believe that a more sophisticated model taking into consideration the use of two hands could show a clearer relationship between QAC engagement and keyboard distance.

Distance from rarest n-gram. We now ask whether users are more likely to engage with QAC when attempting to type words that are difficult to spell. For this analysis we used the rarity of character n-grams as a proxy metric for spelling difficulty, assuming that n-grams that are rare in queries correspond to words that are difficult

[1] http://en.wikipedia.org/wiki/QWERTY

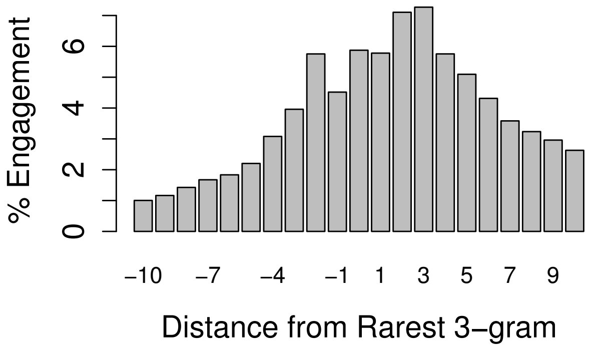

Figure 7: Probability that the user will engage with QAC based on the number of characters they have typed relative to the most difficult to spell section in the word as determined by the position of the rarest 3-gram in the selected query.

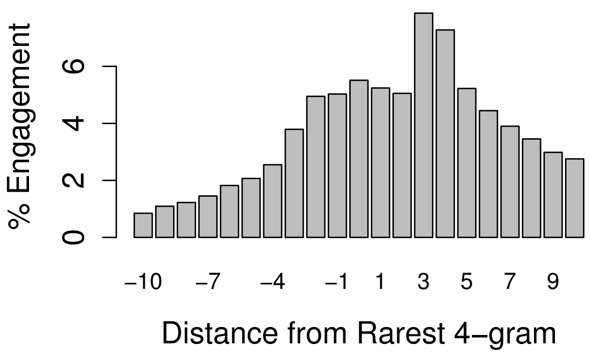

Figure 8: Probability that the user will engage with QAC based on the number of characters they have typed relative to the most difficult to spell section in the word as determined by the position of the rarest 4-gram in the selected query.

to spell. We computed a frequency histogram of all character n-grams of length 3 and 4 based on two years of query logs from Bing. We restricted our n-grams to only those containing characters in the English alphabet. Then, for a given query for which QAC was engaged with, we compute a distance in characters between where the user stopped typing to engage with QAC and the first character of the rarest n-gram in the query. For example, if the user stopped typing after the first character of the rarest n-gram then the distance is 0, and if after the second then 1, and so on.

Figures 7 and 8 show that the highest percentage of QAC engagement correspond to when the user has typed the 3rd character of the rarest tri-gram or the 4-gram in the query. Also there is a steeper decline in percentage of engagement at positions after this point compared to earlier positions. This appears to suggest that users do seem to engage more with QAC leading up to the rarest n-gram, which is likely to be the point at which it becomes clear if the user does or does not know how to spell the word correctly.

4. CONCLUSIONS

In this paper, we investigated user interaction patterns with QAC in Bing. We analyzed auto-completion usage with respect to different attributes of the partial query already typed by the user before engaging with QAC. Our results suggested that users are most likely to engage with auto-completion at word boundaries, when typing rare n-grams, and after typing half of their query characters. We also observed a clear relation between the position of the next character on the keyboard and the likelihood of auto-completion usage. Overall, these results appear to confirm typical intuitions of when users would engage with QAC.

We believe that our results provide valuable insights about how auto-completion is perceived by users. For instance, they suggest that QAC is particularly useful for words that are difficult to spell, and also that engagement with QAC varies with the segment type of the target query. Additionally, the particularly strong engagement at high positions suggests that showing users more suggestions is unlikely to increase engagement, and that ranking the best queries at the top position is critical.

References

[1] J. Arguello, F. Diaz, and J. Callan. Learning to aggregate vertical results into web search results. In *Proc. CIKM*, pages 201–210, Glasgow, UK, 2011.

[2] R. Baeza-Yates, C. Hurtado, and M. Mendoza. Query recommendation using query logs in search engines. In *Proc. EDBT*, pages 588–596, Heraklion, Greece, 2004.

[3] Z. Bar-Yossef and N. Kraus. Context-sensitive query auto-completion. In *Proc. WWW*, pages 107–116, 2011.

[4] M. Ciaramita, V. Murdock, and V. Plachouras. Online learning from click data for sponsored search. In *Proc. WWW*, pages 227–236, Beijing, China, 2008.

[5] K. Grabski and T. Scheffer. Sentence completion. In *Proc. SIGIR*, pages 433–439, 2004.

[6] Z. Guan and E. Cutrell. An eye tracking study of the effect of target rank on web search. In *Proc. CHI*, pages 417–420, 2007.

[7] D. Hawking and K. Griffiths. An enterprise search paradigm based on extended query auto-completion. do we still need search and navigation? In *Proc. ADCS*, 2013.

[8] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. SIGIR*, pages 154–161, 2005.

[9] M. Kamvar and S. Baluja. The role of context in query input: Using contextual signals to complete queries on mobile devices. In *Proc. MobileHCI*, pages 405–412, Singapore, 2007.

[10] E. Kharitonov, C. MacDonald, P. Serdyukov, and I. Ounis. User Model-based Metrics for Offline Query Suggestion Evaluation. In *Proc. SIGIR*, 2013.

[11] M. Shokouhi. Learning to personalize query auto-completion. In *Proc. SIGIR*, pages 103–112, 2013.

[12] M. Shokouhi and K. Radinsky. Time-sensitive query auto-completion. In *Proc. SIGIR*, pages 601–610, 2012.

[13] I. Weber and C. Castillo. The demographics of web search. In *Proc. SIGIR*, pages 523–530, 2010.