

Hey Cortana! Exploring the use cases of a Desktop based Digital Assistant

Rishabh Mehrotra*
University College London
London, United Kingdom
r.mehrotra@cs.ucl.ac.uk

Ahmed Hassan Awadallah², Ahmed El Kholy¹,
Imed Zitouni¹
¹Microsoft, ²Microsoft Research
Redmond, WA, USA
{izitouni,hassanam,ahkhol}@microsoft.com

ABSTRACT

With the advent of improved speech recognition and information retrieval systems, more and more users are increasingly relying on digital assistants to solve their information needs. Intelligent digital assistants on mobile devices and computers, such as Windows Cortana and Apple Siri, provide users with more functionalities than was possible in the traditional web search paradigm. While most user interaction studies have focused on the traditional web search setting; in this work, we instead consider user interactions with digital assistants (e.g. Cortana, Siri) and aim at identifying the differences in user interactions, session characteristics and use cases. To our knowledge, this is one of the first studies investigating the different use cases of user interactions with a desktop based digital assistant. Our analysis reveals that given the conversational nature of user interactions, longer sessions (i.e. sessions with a large number of queries) are more common than they were in the traditional web search paradigm. Exploring the different use cases, we observe that users go beyond general search and use a digital assistant to issue commands, seek instant answers and find local information. Our analysis could inform the design of future support systems capable of proactively understanding user needs and developing enhanced evaluation techniques for developing appropriate metrics for the evaluation of digital assistants.

KEYWORDS

intelligent assistants; user interactions; search sessions; search tasks

ACM Reference format:

Rishabh Mehrotra and Ahmed Hassan Awadallah², Ahmed El Kholy¹, Imed Zitouni¹. 2016. Hey Cortana! Exploring the use cases of a Desktop based Digital Assistant. In *Proceedings of ACM, Tokyo, Japan, August 2017 (CAIR'17)*, 5 pages.

DOI: 10.1145/nmnnnnn.nnnnnnn

*Work conducted while at Microsoft Research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CAIR'17, Tokyo, Japan

© 2017 ACM. 978-x-xxxx-xxxx-x/YY/MM...\$15.00

DOI: 10.1145/nmnnnnn.nnnnnnn

1 INTRODUCTION

Intelligent digital assistants on mobile devices and computers, such as Cortana and Siri, have recently gained considerable attention as increasing number of users interact with them to fulfil their information needs and complete their tasks. Unlike traditional search engines, desktop based digital assistants provide plethora of avenues for users to verbally and textually interact with the system to perform a number of tasks like searching files, setting reminders, surfing the web or even chatting with the assistant. The increased adoption of such novel communication interfaces could partly be attributed to the improved quality of speech recognition as well as the added value users perceive: the spoken dialogue mode of interaction is a more natural way for people to communicate and is often faster than typing. While these novel applications are useful and attractive for users, it is challenging for system designers to understand user interactions with such new interfaces and develop evaluation metrics.

Developing concrete understanding of how the users interact with a digital assistant and the different use cases which help users accomplish their search task provides useful insights in developing user interest models, developing metrics for efficient evaluation of current services for for envisioning future use cases. While studies on traditional information finding tasks on web search abound [TODO: cite], little work has been done to understand user interactions with a digital assistant.

In this work, we present the first use case analysis of a desktop based digital assistant (Windows Cortana) based on large scale logs comprising of over 20 million impressions and 3 million users. We begin by taking a bird's eye view of user interactions by focusing on session level analysis and highlight differences in session boundaries and session characteristics between traditional web search and digital assistant interactions. We then proceed to a more fine grained analysis of user interaction and present four different categories of activity users perform when interacting with Windows Cortana. We present differences in the way users interact with a digital assistant as compared to traditional web search across input methods, use cases, queries and tasks. The analysis presented could help guide development of future user support systems and improve evaluation of current assistants.

2 RELATED WORK

Intelligent assistants possess a unique characteristic as an object of dialogue study. Many dialogue studies have explored the issue of detecting user engagement as well as related affects such as interest and uncertainty [8, 11]. Recently, user satisfaction for intelligent assistants gain attention [2-4]. Jiang et al. [2] proposed an automatic

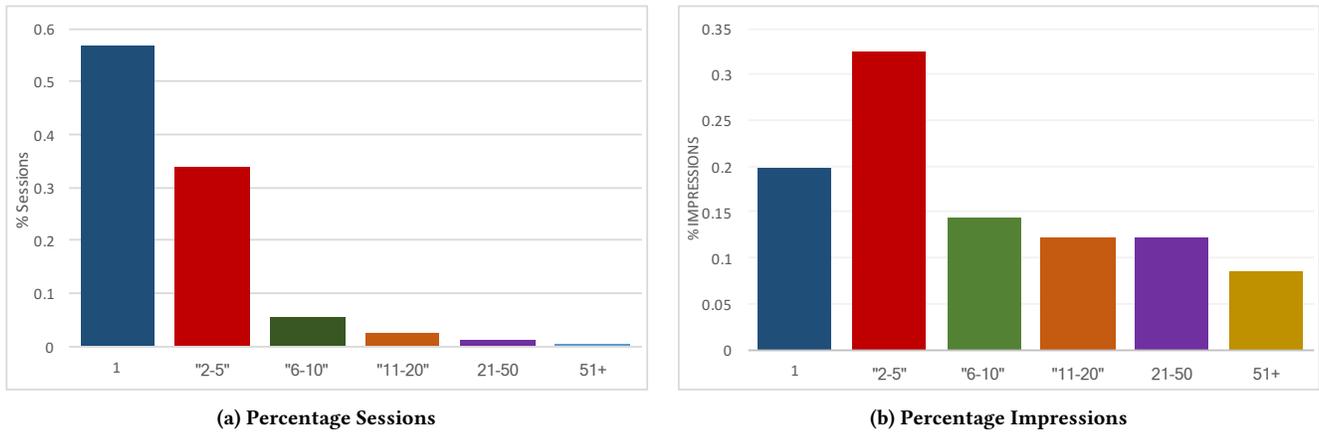


Figure 1: Analyzing session characteristics

method of assessing user satisfaction with intelligent assistants by focusing on simulated tasks for device control, as well as chat and web search, and identifying satisfactory and unsatisfactory sessions based on features used in predicting satisfaction on the web, as well as acoustic features of the spoken request. Kiseleva et al. extended the study of Jiang et al. for detailed understanding [4] and prediction [3] of user satisfaction with intelligent assistants. More broadly, intelligent assistants are often used for longer sessions and tasks that involve sub-tasks and complex interactions, and task complexity has been studied in many user studies. Wildemuth et al. [9] reviewed over a hundred interactive information retrieval studies in terms of task complexity and difficulty, and found that the number of sub-tasks, the number of facets, and the indeterminably were the main dimensions of task complexity. Kobayashi et al. [5] investigated how games played with intelligent assistants affect prospective user engagement. Finally, Shiga et al. [10] describe the creation and examination of over 32K spoken utterances collected during approximately 34 hours of a collaborative search task, based on two dimensions: uncertainty level and need category.

3 ANALYZING USER INTERACTIONS

To better understand the myriad ways in which users interact with a digital assistant, we perform large scale log analysis aimed at extracting insights about user sessions, queries and use cases. In this section, we describe the data context, characterize user sessions and investigate input modularity, query categories and non-traditional use cases.

3.1 Data Context

We analyze a random sample of two weeks of Windows 10 Cortana app logs from June 2016, which contained user interaction data from over 3.6 million users, 21 million impressions and 6 million distinct queries. Each impression referred to one user initiating an interaction with Cortana and was tagged with a user ID and timestamp.

3.2 Session Characteristics

Search sessions have traditionally been considered the focal unit of analysis for seeking behavioral insights from user interactions. In traditional web search setting, it has been reported [1] that more than 79% of the sessions were three or fewer queries with the mean session length being 2.85 queries, with a standard deviation of 4.43. To understand the difference in user interaction between a digital assistant and traditional web search, we begin by investigating session characteristics.

Figures 1a and 1b compare the percentages of sessions and percentage of impressions based on the different session length. We observe that over 80% of sessions have 1-5 queries, which suggests that similar to web search, desktop based digital assistants have a similar length based distribution of sessions. However, as can be observed by looking at impression level aggregates in Fig. 1b unlike traditional web search, longer sessions bring in almost 50% of traffic. Such a significant shift towards longer sessions bringing more traffic hints at more conversational nature of user interactions. In traditional web search setting, long sessions have been seen as occurring when searchers are exploring or learning a new area, or when they are struggling to find relevant information [7]. The prevalence of longer sessions in digital assistant interactions motivates the need for detailed fine-grained analysis on such sessions for better user modelling and metric development.

Session Identification:

The analysis so far assumes session demarcation based on traditional 30 minute activity threshold. However, most prior work has focused on user activity on search engines, insights from which may not be applicable to more novel interfaces like digital assistants. In a recent study [6], we considered the temporal patterns in user initiated interactions and adopted an approach based on Gaussian mixture models to identify session boundary cut-off for user interactions with a digital assistant. We identified an optimal inter-activity threshold for identifying sessions by finding the point where inter-activity time is equally likely to be within the first gaussian fit (within-session) and the second gaussian fit (between-session). The two mixture components fitted correspond

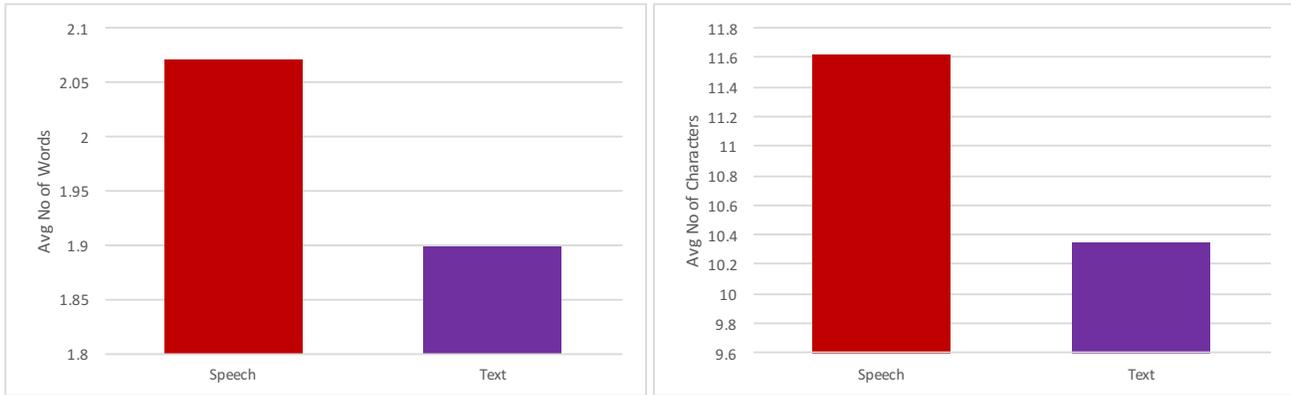


Figure 2: Modality: Speech vs Text

to the within-session interactivity times and the between-session inter-activity times. Our results suggest a smaller threshold (≤ 2 minutes) for session boundary cut-off in digital assistants than the traditionally used 30 minutes threshold for web search engines.

3.3 Characteristic Use-cases & Queries

We next analyze the different use cases and tasks which users perform when interacting with the digital assistant.

3.3.1 Input Modality. While traditional web based retrieval systems are inherently text driven, digital assistants allow users to interact with voice. We analyze differences in query length across the text and speech input modality and present the results in Figure 2. We observe that the query length is more for queries issued via speech commands, and less for manually entered text. The average number of words in a speech query is 2.06, and 1.90 for a text query, with a statistically significant difference of about 10% in the query length in terms of number of words. Similar results are obtained for query length measured in terms of number of characters, with speech queries averaging 11.6 characters, and text based queries averaging 10.3 characters. Further analysis on the query length and session length suggested that longer sessions are mostly speech driven, which hints at the fact that conversational queries are usually longer.

3.3.2 Common Use Case Scenarios. Traditional web search has been about helping users in finding information online and solving their web-based information needs. Novel interfaces of finding information have increasingly made deeper penetration into the market, there by allowing many different non-traditional use cases. In this section, we analyze how different users interact with Windows Cortana to solve different tasks. We categorize user queries into four broad categories:

- (1) **General Search:** The queries falling under this category refer to the traditional web search queries for finding information, e.g. *corporation tax us*.
- (2) **Commands:** In addition to general search queries, users interact with Windows Cortana to issue a number of commands, most of which are system specific commands like alarms, camera, system settings.

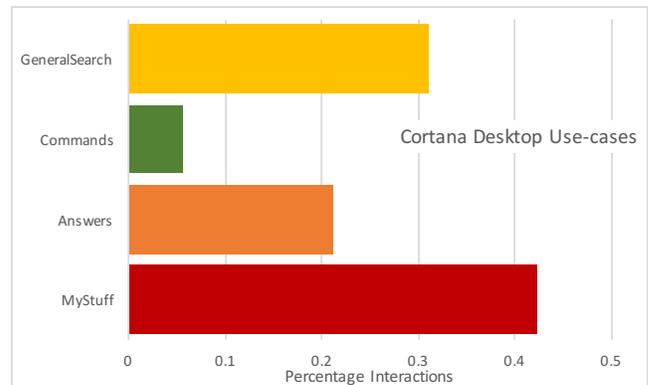


Figure 3: Different Use Cases

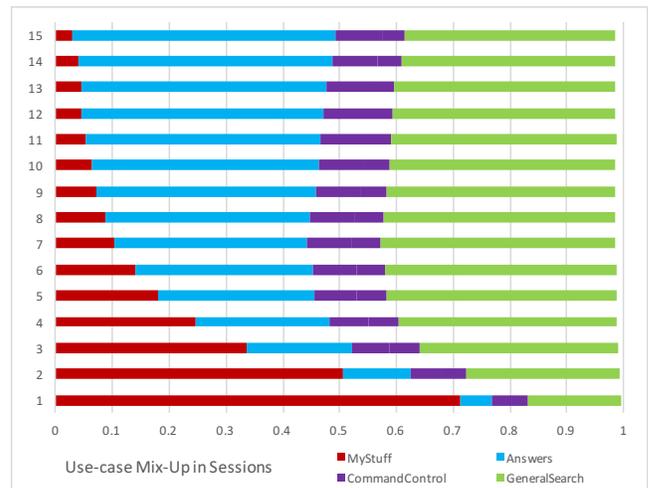


Figure 4: Use-case mix-up in user sessions across different session lengths. The Y-axis is the length of the session in terms of number of queries and X-axis denoting the proportion of sessions of a given length, composed of a particular category.

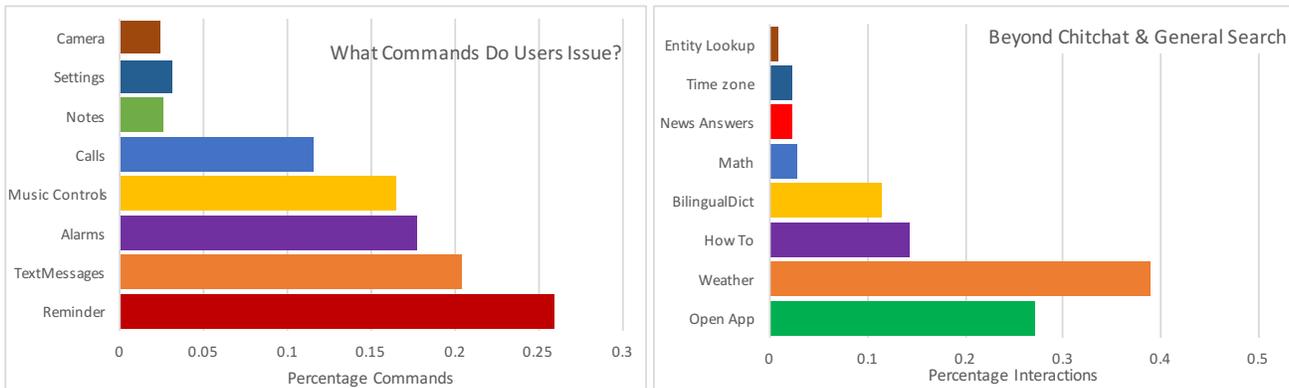


Figure 5: Beyond traditional search: Different types of commands (left) and answers (right) sought by users via interactions with a desktop based digital assistant.

- (3) **Answers:** Given the conversational nature of interactions with digital assistants, often users interact with the aim of finding instant answers. Such interactions and queries fall under this category, e.g. *seattle weather, london time right now*.
- (4) **MyStuff:** A desktop based digital assistant provides a common entry point for users to not only find information over the internet, but also search their local files and folders. Such local information finding queries get categorized into *MyStuff*.

In Figure 3, we show how the user interaction is spread across the above mentioned categories. Indeed, we observe that Cortana desktop is being used for many non-traditional information needs since < 35% of queries belong to the traditional general search paradigm. *MyStuff* dominates the search landscape with over 40% of queries resulting from local information needs. Beyond local search, issuing commands to operate the system received less than 10% of queries. Finally, finding instant answers to queries is increasingly getting more popular, with over 20% of queries resulting in instant answers.

Figure 4 analyzes proportion of user sessions consisting of the above four categories based on the session length. We observe that over 70% single query sessions originate from a local system specific information need (*MyStuff*). The proportion of search sessions concerning local system level information finding steadily decreases as the session length increases. This suggests that finding local machine files and folders are short interactions with the user needing to issue a small number of queries before finding relevant information; unlike web search, wherein user interactions result in longer sessions and more queries. With regard to general web search, we observe a steady spread of sessions across the different session lengths; with similar results holding for *Commands*. The *Answers* category present an interesting case, with an increasing proportion of sessions resulting from instant answer finding queries as the session length increases. Indeed, longer sessions are hypothesized to result from conversational interactions wherein users are trying to find answers in an interactive manner. As conversational systems advance and become better adept at holding conversations,

we believe that more users would rely on such assistants to provide instant answers, thereby increasing the proportion of user queries resulting in instant answers.

3.3.3 Beyond Traditional Search. As shown in the previous section, users use digital assistants in many ways beyond the traditional web search use cases. We next take a deeper look at two specific categories: *Commands* and *Answers* and investigate the different use cases. Figure ?? presents a fine-grained split up of the two categories and shows the top 8 use cases in each. Reminders, alarms and music control are among the most popular commands issued by the users. Given the fact that some users use the digital assistant on their tablet, text messaging and calling also figure among the popular commands. Exploring system settings, taking notes and using the camera constitute other popular commands performed by the assistant. This list highlights the tendency of users in increasingly being more dependent on the digital assistant in performing different functionalities on their device and motivates the need for deeper investigation in understanding and classifying local vs web based information needs.

Figure ?? presents the top-8 instant answers sought by users when interacting with the assistant. Weather related answers and *How Tos* top the list. Similar to traditional web search, Math questions and dictionary usage is common in such assistant usage as well. Finally, quite a few users use the assistant to find news and information about entities of interest. Entities and their accompanying entity panels provide users with concise information which is easily consumed. We envision appropriate changes in these answer categories as the assistant becomes more advanced in tackling more complex scenarios and providing better answers.

4 IMPLICATIONS

The observed differences in the way users interact with a digital assistant compared with traditional web search setting present a number of interesting points for future investigation. First, the prevalence of longer sessions, especially speech based sessions, calls for advancements in interactive and conversational approaches for

information retrieval wherein the system should not only be capable of finding the right information but should also be adept at presenting the user a short verbal summary of results, in order to engage the user in coherent meaningful conversation. Second, as digital assistants become better at more complex tasks, the nature of user interactions and the corresponding sessions characteristics would change. As a result, the session cut-off thresholds would need to be revalidated and the ways session information is used by the assistant would need to be revisited. Third, given the fact that finding instant answers to queries is increasingly getting more popular, with over 20% of queries resulting in instant answers, we envision an increasing tendency of users in being more dependent on the digital assistant in finding precise to-the-point answers. Finally, given the multiple roles these assistants perform in helping users not only do local device specific tasks but also search the web for information, there needs to be deeper investigation in understanding and classifying local vs web based information needs.

REFERENCES

- [1] Bernard J. Jansen, Amanda Spink, Chris Blakely, and Sherry Koshman. 2007. Defining a session on Web search engines. *Journal of the American Society for Information Science and Technology* 58, 6 (2007), 862–871. DOI: <http://dx.doi.org/10.1002/asi.20564>
- [2] Jiepu Jiang, Ahmed Hassan Awadallah, Rosie Jones, Umut Ozertem, Imed Zitouni, Ranjitha Gurunath Kulkarni, and Omar Zia Khan. 2015. Automatic online evaluation of intelligent assistants. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 506–516.
- [3] Julia Kiseleva, Kyle Williams, Ahmed Hassan Awadallah, Aidan C Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Predicting user satisfaction with intelligent assistants. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 45–54.
- [4] Julia Kiseleva, Kyle Williams, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Understanding user satisfaction with intelligent assistants. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. ACM, 121–130.
- [5] Hayato Kobayashi, Kaori Tanio, and Manabu Sassano. 2015. Effects of game on user engagement with spoken dialogue system. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. 422–426.
- [6] Rishabh Mehrotra, Ahmed El Kholy, Imed Zitouni, Milad Shokouhi, and Ahmed Hassan. 2017. Identifying User Sessions in Interactions with Intelligent Digital Assistants. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 821–822.
- [7] Daan Odijk, Ryen W White, Ahmed Hassan Awadallah, and Susan T Dumais. 2015. Struggling and success in web search. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1551–1560.
- [8] Catharine Oertel and Giampiero Salvi. 2013. A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue. In *Proceedings of the 15th ACM on International conference on multimodal interaction*. ACM, 99–106.
- [9] Professor Peter Willett, Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. 2014. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation* 70, 6 (2014), 1118–1140.
- [10] Sosuke Shiga, Hideo Joho, Roi Blanco, Johanne Trippas, and Mark Sanderson. 2017. Modelling Information Needs in Collaborative Search Conversations. (2017).
- [11] William Yang Wang and Julia Hirschberg. 2011. Detecting levels of interest from spoken dialog with multistream prediction feedback and similarity based hierarchical fusion learning. In *Proceedings of the SIGDIAL 2011 Conference*. Association for Computational Linguistics, 152–161.