



What Health Information Are Consumers Seeking? A Comparison Between Two Types of Online Q&A Sites

Ashwag Alasmari^{1,3} and Lina Zhou²

¹ University of Maryland Baltimore County, Baltimore, MD 21250, USA
a.shwag1@umbc.edu

² University of North Carolina at Charlotte, Charlotte, NC 920, USA

³ King Khalid University, Abha, Kingdom of Saudi Arabia

Abstract. Online questioning and answering (Q&A) sites have emerged as an alternative source for serving individuals' health information needs. Despite studies on analyzing user-generated content in online Q&A sites, there is an insufficient understanding of health consumers from the perspective of health-consumer types, information needs and number of questions. Additionally, empirical comparisons of different Q&A platforms are scarce. This research investigates types of health consumers, information seeking needs, and number of questions for asking questions across two types of online Q&A platforms. Empirical analyses of 624 health questions collected from Yahoo! Answers and WebMD reveal several important differences. In comparison, there were more questions about adverse drug reactions on WebMD, and more questions about seeking similar experiences on Yahoo! Answers. The findings have design implications for online Q&A sites to better support health information seeking.

Keywords: Information needs · Online health seeking · Online Q&A

1 Introduction

Health consumers are increasingly going online to access health information. According to the 2014 Pew Internet and American Life Project survey, 72% of Internet users reported using online resources to obtain health information in the past 12 months [5]. Specifically, over 35% of respondents searched online for medical information to determine what medical condition they or someone else might have. These online resources ranged from general search engines to specific websites devoted to health information. One of the many online resources that meet these information needs is online Question and Answer (or Q&A) sites.

Online Q&A sites [21] allow health consumers to post questions with expectation of getting answers from others. Accordingly, such sites represent the most direct way for a person to seek information. Q&A sites also serve as an alternative source to search engine after consumers fail to find information via the latter [9]. In Q&A research, health has been identified as a major domain for observing user interactions [12]. There are different types of online Q&A sites. Some sites such as Yahoo! Answers are

organized as a directory with different sections focusing on different topics, whereas some others like Quora do not provide pre-defined categories for questions. Online Q&A sites also differ based on whether the posted answers are curated or not. In a curated site such as WebMD Answers, experts are considered as an essential part of the community. In contrast, health consumers are able to seek support from peers with similar conditions in some other online Q&A platforms. However, none of the previous studies have compared health consumers' needs across different types of Q&A sites.

To fill the above literature gaps, this research compares curated and community Q&A sites. Specially, it answers the following research questions: (1) How do the health information needs differ between the two online Q&A sites, if any? (2) Are there differences between different types of Q&A sites in terms of health-consumer types and number of questions?

The study answers the above questions by analyzing health questions collected from two distinct online Q&A platforms: Yahoo! Answers and WebMD Answers. According to the statistic of National Institute of Diabetes and Digestive and Kidney Diseases, kidney disease can lead many complications, and the overall prevalence of chronic kidney disease in general population estimated to be 14%. Thus, data are collected pertaining to questions about the kidney disease from the two platforms in this study.

2 Background and Literature Review

Online Q&A sites provide a venue generally for asking a question and posting answers. While online Q&A usually refers to user-generated answers, there are examples of systems that do automatic extractions of answers, such as Google or Bing, which are also known as machine-driven Q&A. On the other hand, online Q&A has evolved owing to the significant development of Web 2.0. It is an outlet for information seeking where the health consumers' needs are articulated by natural language questions and posted to a community. Therefore, answers to the questions are answered by anyone who share a topic of interest from the community. The answers can include, but are not limited to, information, social, suggestions, advice, or opinions.

With the growing popularity of online Q&A services in recent years, many classification schemes have been proposed for such Q&A sites. Q&A platforms can be focused on a specific topic or general topics. Examples of specific or specialized Q&A include WebMD and Stack Overflow. There is an abundance of examples which fall under the general-purpose Q&A category. These sites typically cover a broad range of topics instead of being organized around just one single (broad) topic. Examples of general focused Q&A include Yahoo! Answers and Quora. To make this classification more precise, Q&A sites further differ based on whether the posted answers are curated or not. In a curated site, experts are considered as an essential part of the community such as WebMD Answers and Quora. Take WebMD as an example, answers posted by experts are featured and ranked top on WebMD Answers, followed by answers posted by other types of contributors. In contrast, health consumers are able to seek support from peers as well in some other Q&A sites such as Yahoo! Answers and Quora and the answers are featured based on the total number of community votes. Choi et al. [3]

proposed a typology of online Q&A platforms consisting of four categories: community-based (such as Yahoo! Answers) Q&A, collaborative Q&A (such as WikiAnswer), expert-based Q&A or curated-based Q&A (such as WebMD Answers) and social Q&A (such as Q&A hashtags on Twitter).

Online Q&A sites provide an alternative paradigm for seeking and sharing information to search engines [21]. These type of sites not only avoid dealing with a large number of search results before getting at the desired information, but also help directly find related information in a short period of time based on the information given in questions. Additionally, the answers are given by other users with knowledge or similar experiences, which are likely to be useful and easy to understand to the information consumers [10]. One promising application of online Q&A is in health, such as health-related information seeking.

The extant studies on online Q&A in health-related topics have centered on two issues: (1) content [15, 23] (e.g., questions and answers) and (2) community dynamics [4] (e.g., information questioners, answerers, and the community in general). Information needs and consumers' behavior in online Q&A have been analyzed for various medical conditions such as HIV/AIDS [14], cancer [16], STD [13], diabetes [22] and H1N1 [7]. One major stream of research in the area of information seeking behavior is comparing and contrasting different Q&A sites. A recent study explored how the health information seeking behavior of lay persons differs from that of professionals across 5 different sites [18]. Based on an analysis of linguistic characteristics of consumers' needs (such as lexical, syntactic, and semantic information), they found that patients ask longer questions, provide background information, and ask for different types of information compared to professionals. Another study aimed to understand health consumers' usage of medical concepts by evaluating the coverage of concepts and semantic types of the Unified Medical Language System (UMLS) in two types of social media: blogs and Q&A [17]. The findings reveal that UMLS concepts appeared more frequently in social Q&A postings when compared to blog postings.

The review of related literature suggests that there is a lack of understanding of possible differences concerning health questions between community and curated online Q&A sites. In addition, they fail to understand the health consumers from the perspective of health-consumer types, information needs and the number of questions. This research aims to address the above limitations by comparing health questions across two main online questioning and answering (Q&A) platforms: community and curated Q&A sites. The comparison is performed along three main dimensions: health-consumer types, health information seeking needs, and the number of questions.

3 Methods

3.1 Q&A Sites Selection

We selected Yahoo! Answers as a community-driven Q&A site for data collection because health is one of the top-level categories in this platform. The platform was released in December 2005, becoming the third most popular internet reference site in the world and the most visited community Q&A site in the United States; 16.64% of

Yahoo! users used Yahoo! Answers [24]. There is a diversity of topics in which health consumers can participate in. For the purpose of our research, we collected 9 years' worth of data (2006 – 2015) from the site.

We chose WebMD Answers as a Curated Q&A site. As one of the most influential online health Q&A websites [2, 11], WebMD is served by professional health organizations, and American certified health experts in a broad range of specialty areas and registered site users. The site covers over 900 health topics ranging from acne to weight loss, and it also serves as a platform for people with similar health concerns and wellness interests to meet and share experiences as well as interact with certified health experts and specialists. We collected all posts starting at August 2008 from the site.

As explained in Introduction, we selected questions related to a specific disease condition of the kidney in this study. The screening of related questions was based on related search keywords or key-phrases such as *kidney*, *kidney infection*, *kidney stone*, *kidney cancer*, *kidney disease*, *chronic kidney disease*, *dialysis*, *kidney failure*, *renal artery stenosis*, and *renal cell carcinoma*. For the purpose of this study, we randomly sampled 316 questions from Yahoo! Answers across various kidney related topics by using the stratified sampling method. In addition, we randomly selected 308 questions from WebMD.

We cleaned the datasets by removing noise such as advertisements, non-human subjects (e.g., questions about the kidney problems of pets instead of humans), questions that were non-kidney related, questions related to a student's projects, and posts that did not actually state any questions. The final dataset consisted of 624 questions from the two platforms.

3.2 Dataset Description

For each question post, we collected related information such as questions titles, and descriptions, post-date, and answers. In addition, we collected some other information about the questions such as the categories where the questions posted, total number of answers and the number of votes for each answer. In this study, we only used question titles and descriptions. The descriptive statistics of datasets are reported in Table 1.

Table 1. Descriptive statistics of the Yahoo! Answers and WebMD answers datasets

Yahoo! Answers			WebMD		
Category	# Q	%Q	Category	# Q	%Q
Other diseases	218	69	Kidney	162	52.6
Diabetes	21	6.6	Kidney infection	28	9.1
Heart diseases	8	2.5	Kidney stone	72	23.8
Infectious diseases	10	3.2	Kidney cancer	10	3.24
Cancer	6	1.9	Kidney disease	24	7.8
Other health related topics	36	11.4	Renal cell carcinoma	2	0.60
General topics	17	5.8	Renal artery stenosis	10	3.24

3.3 Analysis Method

Content analysis is one of the most widely used research methods for analyzing the questions and answers in online Q&A platforms [12]. During the content analysis, questions are coded into a set of categories and these categories and frequencies are used to understand the dataset. More specifically, we used directed content analysis in which we followed predefined categories from a previous study [23].

We developed an annotation system to facilitate the analysis of questions. Two coders analyzed the questions independently. The questions are randomly assigned to two coders at a ratio of 6:4. To check the validity of the coding results, we randomly drew a sample of 20% of questions, which were then reanalyzed by a third coder independently. The inter-coder agreement of the sample set was 87.9%. We then discussed discrepancies with the coders to eliminate potential biases.

3.4 Analysis Categories

To answer the research questions, we focused on the following types of information in content analysis.

Health-Consumer Types. A health consumer may post questions about him/herself or on behalf of someone else. Accordingly, the questions were labeled as self and other, respectively. In the latter case, the health information described in the question was typically about the health consumers' family member or friend. In addition, we used non-identified as a third value to cover questions that did not have a clear indication of the role of the health consumers.

Information Seeking Needs. Health consumers ask health-related questions online to address their specific needs. The following is a list of common health information needs in asking questions in online Q&A sites, introduced in a previous study [23].

- Symptom: to gain an understanding of symptoms of a kidney or any other related disease.
- Diagnosis: to confirm the nature of certain disease.
- Causes: to figure out the causes of the disease.
- Prognoses: to inquire about the hypothetical effect of a disease.
- Treatment: to explore treatment alternatives of a kidney disease.
- Supplements and lifestyle: to explore lifestyle and diet in people with a kidney disease as well as using different supplements.
- Information sources, medical profession, and related types of information: to look for medical experts in the field and any kind of resources to fulfill health consumers' needs.
- Drug interaction: to ask for more details about unfavorable and unexpected signs, symptoms, or diseases associated with the use of a drug without any judgment about the causality or relationship to the drug use.
- Similar experiences: to connect to patients with similar conditions.

Number of Questions. The number of sub-questions per question, including the title is calculated in this study. Some questions are asked in different formats; we can only consider one question that shows the same information needs.

4 Results

To fully answer the research questions, we conducted independent sample t-test in addition to reporting descriptive statistics.

4.1 Health-Consumer Types

Table 2 shows the distributions of health-consumer types of the two selected Q&A sites. Overall, more than 60% of health consumers asked questions about themselves or others across the websites. Nevertheless, more questions were posted by family members or friends on behalf of patients on Yahoo! Answers (19.9%) than on WebMD (9.7%). On the other hand, there is a significant percentage of health consumers whose types were not identified, which accounted for 37.7% questions from Yahoo! Answer and 41.2% from WebMD, respectively.

Table 2. Distribution of health-consumer types

Types	Yahoo! Answers	WebMD
	Distribution	Distribution
Self	42.4%	49%
Other	19.9%	9.7%
Non-identified	37.7%	41.2%

4.2 Information Seeking Needs

We identified 10 types of information needs of health consumers, including diagnose, cause, treatments, supplements and lifestyle, prognoses, symptoms, information sources, drug interaction, similar experiences, and others. The descriptive statistics of the different types of information needs (in terms of percentage) is summarized in Table 3. In addition, we report the statistical test results of independent sample t-test in the last column.

Table 3. Descriptive statistics and t-test results of information seeking goals

Information needs	Yahoo! Answers		WebMD		p-value
	Mean	SD	Mean	SD	
Diagnose	0.24	0.43	0.18	0.39	0.0725'
Cause	0.19	0.39	0.19	0.39	0.958
Treatment	0.21	0.4	0.15	0.36	0.0837'
Supplements and lifestyle	0.15	0.35	0.13	0.34	0.6539
Prognoses	0.18	0.38	0.09	0.29	0.0011**
Symptom	0.15	0.36	0.08	0.27	0.0036**
Information source	0.08	0.27	0.07	0.26	0.952
Drug interaction	0.02	0.14	0.07	0.25	0.0027**
Similar experiences	0.07	0.26	0	0	1.09E-06'
Other	0.18	0.38	0.09	0.29	0.0023**

Note: *: significant at .01, ': significant at .1

The most common information need of using both Q&A sites was about diagnosis of a health condition (21.2%) (e.g., *"I have been researching Chronic Kidney Disease on WedMd and I've had some of the [symptoms]. Such as lots of headaches, not feel hungry, nauseated.....I've been to the doctors twice. Only one of them a urine sample was taken. ...Anyone know anything useful."*)

The next two common information needs are inquiring about causes (18.6%) and treatment (17.9%). For example, *"How does chronic kidney disease affect homeostatic mechanisms? What is the cause and what is the effect on homeostatic mechanisms?"* *"What is pentoxifylline? [Can] this can help for chronic kidney disease?"*. Questions in the supplements and lifestyle (diet and exercise) category often asked for recommendations on diets or food should be taken to help stay healthy, or recover from a disease. For example, *"is there treatment to cure it? and also what foods to avoid."*

Prognoses is another common information need for questioning. For example, someone asked a question about his father who was recently diagnosed with CKD and had only 30% of his kidney functioning: *"how much longer will he live? how much longer till he needs dialysis/transplant?"*

For questions focusing on symptoms (11.5%) are relatively general. For example, *"Does anyone know the signs of chronic kidney disease?"* and *"How can you tell if you have chronic kidney disease[?]"*

Some health consumers visited the Q&A sites for recommendations on information sources (7.5%). Among the questions, 4.3% were motivated by understanding medication effects, such as learning about side effects (e.g., *"is there a blood pressure medicine that [I] can take that will not harm kidneys"*), comparing and contrasting medications (e.g., *"Does Renal interact with any other medications?"*), or seeking information on available medications, such as whether non-prescription options are available or learning more about medications and interaction between different medications.

Beyond these broad categories, health consumers also asked for advice and experiences from those who were experiencing similar conditions (3.7%). One example is that a

health consumer shared her story about some symptoms and was not sure if the hair loss was related to kidney disease. At the end of the question, she asked, “*Does anyone know of anyone that has [experienced] hair loss [due] to a failed kidney or has experienced it themselves?*” In addition, another health consumer shared her husband’s case who had recently been diagnosed with CKD stage III and she was wondering about the progression to dialysis. She emphasized at the end of her post that “*Just looking for someone with personal experience.*” Another patient who was in early 30s was diagnosed with CKD. He stated, “*The doctor is saying that I am developing chronic kidney disease. Has anyone else been diagnosed with this so young? If so, what lifestyle changes have you had to do to help combat it?*”

The results of T-tests revealed some significant differences in the health information needs between the two platforms. For those needs that showed differences, most of them are more common in Yahoo! than WebMD Answers, such as symptoms, prognoses, similar experience ($p < .01$), and diagnoses and treatment ($p < .1$). The only exception was adverse drug reaction, which were more frequently asked in WebMD ($M = 0.07$, $SD = 0.25$) than Yahoo!! Answer ($M = 0.02$, $SD = 0.14$) ($p < .001$).

4.3 Number of Questions

The distributions of number questions are reported in Table 4. The data shows that the majority of questions (92.5%) posted on WebMD Answers contains only one question. In contrast, more than 40% of questions on Yahoo! Answers include multiple questions.

Table 4. Distribution of health-consumer types

# of questions	Yahoo! Answers	WebMD
1	57.9%	92.5%
2	31.1%	21.1%
3	7.2%	1.9%
>= 4	3.8%	0

5 Discussion

Q&A platforms generally provide a venue for asking a question and posting answers [12]. Accordingly, such sites represent the most natural way for a person to seek information by supporting the writing of questions fully and explicitly. This research investigates and compares health-consumer types, health information needs and number of questions across two types of online Q&A platforms. Empirical analyses of 624 health questions collected from Yahoo! Answers and WebMD Answers reveal several important differences.

First, more questions in general were posted by family members on Yahoo! Answers than WebMD. The process of posting a question on the former platform is relatively simpler than the latter. Yahoo! Answer only requires the email address of a health consumer, whereas WebMD not only requires more information, but also encourages the health consumer to create a profile and provide more details about his/her health status.

One main difference between the community-driven and curated online Q&A platforms lies in the type of health information needs. There were more questions concerning adverse drug reactions posted on WebMD than on Yahoo! Answers. A curated site carries authoritative expertise with it, and topics like adverse drug reactions and drug-to-drug interactions are complicated, and responding to such questions requires extensive domain expertise. Although WebMD allows other health consumers to provide answers in addition to medical professionals, the responses from the latter group were given more weights and top listed. This finding has implications for discovering new adverse drug reactions from user-generated content online [1, 8, 20]. To this end, WebMD can serve as a more appropriate source than Yahoo! Answers.

In contrast, a significant percentage of questions posted on Yahoo! Answers were aimed to seek answers from peers with similar conditions, which were absent in WebMD. Community Q&A is a social interaction platforms and built based on user-generated content and the interaction among users. Yahoo! Answers is a large and diverse Q&A platform, acting not only as a medium for sharing technical knowledge, but as a place where one can seek advice, gather opinions, and satisfy health consumers' curiosity about a variety of topics [12]. They include sharing personal stories and experiences, among others, which is confirmed in our study. By the same token, the findings of this study reveal that health consumers were more likely to ask multiple questions at Yahoo! Answers more than WebMD.

Design Implications

The themes emerging from our findings have implications for improving online Q&A sites.

It is important for Q&A sites to support social interaction for effective communication of medical conditions. Our results reveal that family members were active in seeking health information on behalf of patients on Yahoo! Answers. The literature has shown that many people with chronic and unstable conditions co-manage or delegate health management activities to others [19]. However, family members and close friends are often not aware of patients' specific health conditions, treatment recommendations, or care goals, and consequently may not know how to best support lifestyle and medication treatment regimen adherence or decision-making preferences [6]. Therefore, providing support for social interaction would help improve the accuracy and completeness of information described in question posts on these Q&A sites. In addition, extending online Q&A platforms to include family members and friends can facilitate more proactive and impactful involvement of family members in patients' healthcare self-management.

It is worth noting that many health consumers value peers' personal experience and specifically sought for people whom they can talk to. Online Q&A has emerged as a popular and effective paradigm for meeting a wide range of information needs [12]. There is an enormous amount of knowledge and expertise shared in online Q&A sites. As the number of questions posted about looking for similar patients in online Q&A sites increases, routing these questions to anyone who share similar experiences is another way of direction. This is expected to contribute to improved efficiency, engagement, and survivability of online Q&A community. In addition, identifying "authoritative" users and promoting their posts would help better meet the health

information needs of other users. Further, online recommendation techniques can be adapted to suggest other related questions and responses to a user's question post in online Q&A sites.

Future Directions

Like all other studies on online information seeking behavior, this study has some limitations. First, we collected data from two online Q&A sites only, and thus the findings may not be generalizable to other online health Q&A platforms. This study can be extended by examining questions about different type of diseases and from other types of platforms. Second, despite that the analysis of question content helped gain insights into contextual characteristics of questions, other types of data from the online Q&A sites such as user activities, profiles, comments, and ratings are also valuable. An integration of data from the various dimensions will help gain a fuller understanding of user behavior in online Q&A communities.

6 Conclusion

Through analyzing questions collected from community Q&A and curated Q&A sites, this study identified both of their similarities and differences in term of health-consumer type, information seeking needs and the number of questions. The findings have implications for the design of online Q&A websites and for effective use of such platforms to meet patients' health information needs.

References

1. Benton, A., Ungar, L., Hill, S., et al.: Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. *J. Biomed. Inf.* **44**(6), 989–996 (2011)
2. Bramall, A.N., Bernstein, M.: Improving information provision for neurosurgical patients: a qualitative study. *Can. J. Neurol. Sci./Journal Canadien des Sciences Neurologiques* **41**(01), 66–73 (2014)
3. Choi, E., Kitzie, V., Shah, C.: Developing a typology of online Q&A models and recommending the right model for each question type. *Proc. ASIST Ann. Meeting* **49**, 1 (2012)
4. Choi, E., Scott, C.R.: Effects of user identity information on key answer outcomes in social Q&A. In: *iConference*, pp. 302–315 (2013)
5. Susannah Fox. The social life of health information (2014). <http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/>
6. Kernisan, L.P., Sudore, R.L., Knight, S.J.: Information-seeking at a caregiving website: a qualitative analysis. *J. Med. Internet Res.* **12**, 3 (2010)
7. Kim, S., Pinkerton, T., Ganesh, N.: Assessment of H1N1 questions and answers posted on the web. *Am. J. Infect. Control* **40**(3), 211–217 (2012)
8. Lardon, J., Abdellaoui, R., Bellet, F., et al.: Adverse drug reaction identification and extraction in social media: a scoping review. *J. Med. Internet Res.* **17**(7), e171–e171 (2015)
9. Liu, Q., Agichtein, E., Dror, G., Maarek, Y., Szpektor, I.: When web search fails, searchers become askers. In: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR 2012*, p. 801 (2012)

10. Molino, P., Aiello, L.M., Lops, P.: Social question answering: textual, user, and network features for best answer prediction. *ACM Trans. Inf. Syst.* **35**(1), 4–40 (2016)
11. Nowrouzi, B., Gohar, B., Nowrouzi-Kia, B., Garbaczewska, M., Brewster, K.: An examination of scope, completeness, credibility, and readability of health, medical, and nutritional information on the internet: a comparative study of wikipedia, WebMD, and the mayo clinic websites. *Can. J. Diabetes* **39**(2015), S71 (2015)
12. Oh, S.: Social Q&A. In: *Social Information Access*, pp. 75–107 (2018)
13. Oh, S., Park, M.S.: Text Mining as a Method of Analyzing Health Questions in Social Q & A, pp. 1–4 (2013)
14. Oh, S., Park, M.S.: HIV/AIDS Question Analysis with Text Mining : Using Concept Maps for Data Analysis and Interpretation Results & Discussion (2014)
15. Sanghee, O., Worrall, A.: Health answer quality evaluation by librarians, nurses, and users in social Q&A. *Libr. Inf. Sci. Res.* **35**(4), 288–298 (2013)
16. Sanghee, O., Zhang, Y., Park, M.S.: Cancer information seeking in social question and answer services: identifying health-related topics in cancer questions on Yahoo!! answers. *Inf. Res. Int. Electron. J.* **21**, 3 (2016)
17. Park, M.S., He, Z., Chen, Z., Sanghee, O., Bian, J.: Consumers’ use of UMLS concepts on social media: diabetes-related textual data analysis in blog and social Q&A sites. *JMIR Med. Inf.* **4**(4), e41 (2016)
18. Roberts, K., Demner-Fushman, D.: Interactive use of online health resources: a comparison of consumer and professional questions. *J. Am. Med. Inf. Assoc.* **23**(4), 802–811 (2016)
19. Ann-Marie Rosland, M.P.A., Michele Heisler, M.D., Janevic, M.R., et al.: NIH public access. *Fam. Syst. Health* **31**(2), 119–131 (2013)
20. Sarker, A., Ginn, R., Nikfarjam, A., et al.: Utilizing social media data for pharmacovigilance: a review. *J. Biomed. Inf.* **54**, 202–212 (2015)
21. Shah, C., Oh, S., Oh, J.S.: Research agenda for social Q&A. *Libr. Inf. Sci. Res.* **31**(4), 205–209 (2009)
22. Zhang, J., Zhao, Y., Dimitroff, A.: A study on health care consumers’ diabetes term usage across identified categories. *Aslib J. Inf. Manage.* **66**(4), 443–463 (2014)
23. Zhang, Y.: Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In: *Proceedings of the 1st ACM International Health*, pp. 210–219 (2010)
24. The top 500 sites on the web. <https://www.alexa.com/topsites/category/Top/Reference>. Accessed 20 Sept 2018