Exploring People's Attitudes and Behaviors Toward Careful Information Seeking in Web Search

Takehiro Yamamoto
Kyoto University
Kyoto, Japan
tyamamot@dl.kuis.kyoto-u.ac.jp

Yusuke Yamamoto
Shizuoka University
Hamamatsu, Japan
yamamoto@inf.shizuoka.ac.jp

Sumio Fujita
Yahoo Japan Corporation
Tokyo, Japan
sufujita@yahoo-corp.jp

ABSTRACT

This study investigates how people carefully search for the Web to obtain credible and accurate information. The goal of this study is to better understand people's *attitudes* toward careful information seeking via Web search, and the relationship between such attitudes and their daily search *behaviors*. To this end, we conducted two experiments. We first administrated an online questionnaire to investigate how people's attitudes toward using the strategies for verifying information in the Web search process differ based on various factors such as their credulity toward Web information, individual thinking styles, educational background, and search expertise. We then analyzed their one-year and one-month query logs of a commercial Web search engine to explore how their daily search behaviors are different according to their attitudes.

The analysis of the questionnaire and the query logs obtained from 1,491 participants revealed that (i) the people's attitudes toward using the verification strategies in Web search are positively correlated to their Need for Cognition (NFC), educational background, and search expertise; (ii) people with strong attitudes are likely to click lower-ranked search results than those with intermediate levels of attitude; (iii) people with strong attitudes are more likely to use the terms such as "evidence" or "truth" in their queries, possibly to scrutinize the uncertain or incredible information; and (iv) the behavioral differences found in (ii) and (iii) are not identified from the differences in the participants' educational backgrounds. These findings help us explore future directions for a new Web search system that encourages people to be more careful in Web search, and suggest the need for an educational program or training to facilitate the attitudes and skills for using Web search engines to obtain accurate information.

CCS CONCEPTS

• **Information systems → Web searching and information discovery**;

KEYWORDS

Web Search, Behavior Analysis, Credibility

1 INTRODUCTION

While Web search is an important tool for people access information on the Web to make various decisions in their daily lives, the lack of credibility or the dissemination of the inaccurate information on the Web have been crucial problems in our society. For example, Sillence *et al.* reported that more than the half of existing medicine-related Web pages have not been authorized by any medical experts [20]. White also revealed that, for more than the half of the medical "yes–no" queries, their top-ranked search results provided incorrect answers according to the physicians' evaluations [23].

However, several studies have revealed that many people regard the information on the Web as somewhat credible and may not be aware of such problems. According to a survey conducted by Nakamura *et al.* [18], 57% participants answered that they trusted the information returned by a Web search engine. A more recent survey [13] also reported that about 70% of the respondents from Japan were not likely to question the credibility of Web information such as online news articles or product reviews.

To obtain the credible and accurate information during a Web search process, people need to use several *strategies* to verify information, such as issuing multiple queries to collect information from diverse perspectives, or actively checking whether the information on the Web page is up-to-date. While a few studies have attempted to better understand the extend to which people use such strategies based on a questionnaire [16–18, 32], such studies are generally limited. We could understand people's *attitudes* toward using such strategies from the analyses [17, 32], but the behavior of people in routine Web search processes remains unclear. The goal of the present study is to better understand the degree to which people are careful about using Web search engines and examining the information retrieved by them. More specifically, we aim to address the following two research questions:

RQ1 Which factors do affect people's attitudes toward using verification strategies in Web search?

RQ2 Are there differences in people's daily Web search behaviors based on their attitudes? If yes, what are these differences?

To address these research questions, we first administer an online questionnaire to investigate how people's attitudes toward using the verification strategies differ based on various factors such as people's credulity toward Web information, individual thinking styles, educational background, and search expertise. We then analyze their one-year and one-month query logs of a commercial Web search engine to explore how their daily search behaviors differ according to their attitudes. We expect that these two analyses will help us find the relationship between people's attitudes and their actual Web search behaviors and design a Web search system that encourages people to obtain credible and accurate information.

The analysis of the questionnaire and the query logs obtained from 1,491 participants revealed the following: (i) the people's attitudes toward using verification strategies in Web search are positively correlated to their Need for Cognition (NFC), educational backgrounds, and search expertise; (ii) people with strong attitudes are likely to click lower-ranked search results than people ranked intermediate on attitude. (iii) people with strong attitudes are more likely to use terms such as "evidence" or "truth" in their queries, which might be intended to scrutinize the uncertain or incredible information; and (iv) the behavioral differences found in (ii) and (iii) are not identified from the differences in their educational backgrounds. These findings will help us explore future directions for a new Web search system that encourages people to be more careful with Web search, and suggest the need for an educational program or training to facilitate the attitudes and skills of using Web search engines to obtain accurate information.

2 RELATED WORK

2.1 User Characteristics and Search Behavior

People's Web search behaviors vary based on different characteristics. Hölscher and Gerhard investigated the effect of search expertise on people's Web search behavior [11]. They reported that the search experts tend to use more advanced search options than the others to formulate queries. White and Morris [26] also analyzed the difference between search experts and the others based on their query logs. They found that the search experts, who use the advanced query operators, are more likely to query less frequently in a session, compose longer queries and success in search. Domain knowledge has also been a topic of interest of the research community [4, 10, 25]. Domain experts can formulate more complex queries [10] and are more likely to succeed at Web search [25]. Several researchers have attempted to understand the gender difference in the perceptions and behaviors on Web search [14, 34]. For example, Zhou investigated how gender affects the task performance [34]. Weber and Jaimes analyzed query logs of users on a large-scale, relating them to the U.S. census data [22]. They revealed the relationship between user demographics such as gender, age, educational background, income, and their search behaviors such as dwell time, the number of clicks, and topics of search.

2.2 Web Information Credibility

Many researchers have proposed systems to support users in judging the credibility of Web information. Yamamoto et al. proposed a system that provides the criteria for credibility judgment such as authority, currency, and topical coverage of each search result on search engine results pages [30]. A few researchers have proposed methods to show disputed sentences/topics during Web search and browsing [6, 29] so that people can be made aware the existence of the opposing opinions on the Web. Yamamoto et al. proposed a query suggestion technique that utilizes the priming effect, which is a well-known psychological effect, to enhance the people's attitudes toward critical thinking [31].

Understanding people's perceptions of the credibility of Web information is another important research direction. Nakamura et al. conducted an online survey with 1,000 participants to understand how and when they use the Web search engine, and how they rely on the search results returned by a Web search engine [18]. Metzger et al. conducted questionnaire-based user studies on how college students perceived the information on the Web and the types of verification strategies employed by them [17]. The part of the questionnaire items used in our online questionnaire is adapted from their study. Metzger et al. further conducted a study with more than 2,000 young students to understand the relationships among their perceptions of the credibility of Web information, personality variables such as age, academic performance, thinking styles, and the skills for evaluating Web information [16].

3 CAREFUL INFORMATION SEEKING IN WEB SEARCH

We first introduce desirable behaviors of people for obtaining credible and accurate information via Web search, and explain the attitudes to be investigated in the present study. We then discuss the possible factors that influence such attitudes from the literature, and propose the hypotheses to be investigated in the questionnaire analysis. Finally, we discuss the possible behavioral differences we expect between people ranked high on attitude and the other people and propose the hypotheses to be investigated in the query log analysis.

3.1 Attitude Toward Using Verification Strategies in Web Search

To obtain accurate information via Web search, people are required to properly use a Web search engine and critically examine the information they encounter. Such skills and attitudes are closely related to people's critical thinking and information literacy skills. According to Ennis, critical thinking is a logical and reflective style of thinking to determine what to believe or do [7]. Information literacy refers to the ability to identify information needs, find the required information efficiently, evaluate the information critically, and use it [12].

Many researchers have pointed out that people need to expend efforts on verifying the credibility and accuracy of online information in terms of accuracy, authority, objectivity, currency, and coverage [17, 30]. Metzger et al. proposed several verification strategies for obtaining accurate online information and conducted a questionnaire-based survey to understand the extent to which young students use such strategies, including the following [17]:

- I try to check whether the information on the Web page is up-to-date.
- I try to check whether the information on the Web page is complete and comprehensive.

(See Section 4.2.2 and Table 2 for the details about such strategies).

We regard that the people who answered that they are willing to expend efforts on those verification strategies in Web search to obtain credible and accurate information as people ranked high on *attitude* toward using verification strategies in Web search. We refer to people's attitudes toward using the verification strategies in Web search as *verification attitudes* hereinafter. Since the responses to the questionnaire are based on the participants' subjective experiences, and they may contain social desirability response bias, we further analyzed the participant's daily Web searching logs to understand how people's verification attitudes affect their search *behaviors*.

3.2 Factors Affecting Verification Attitudes

To address the research question **RQ1** described in Section 1, we discuss the possible factors that would contribute to knowing what kind of people are likely to have strong verification attitudes.

3.2.1 Credulity toward Web information and search results. A person's perception of the credibility of the Web Information is a fundamental factor affecting how they carefully examine the information they find via a Web search. A person who doubts the credibility of the information on the Web in general is expected to carefully examine the results, while a person who believes in the information on the Web is less likely to employ the verification strategies in Web search.

In addition to the perception of the credibility of the Web Information, we also consider the credulity toward the search results returned by a search engine. Some study revealed that the people's believability of the search results varies by individuals [18]. Also, many studies have revealed the existence of a position bias in the ranking of search results; people are more likely to click the higher-ranked results beyond their relevance [33]. A person who trusts the search results or the top-ranked search results may not try to implement any strategy for verifying the results. Based on these discussions, the following hypothesis is investigated in the present study:

> **H1-1** The credulity toward Web information and search results is negatively correlated to the verification attitude.

3.2.2 Thinking styles. Individual differences in thinking styles, that is, how people think or evaluate information in general, may affect their verification attitudes. For example, Need for Cognition (NFC) [3], which reflects how much a person enjoys effortful cognitive activities, is one of well-known thinking styles. For example, Metzger *et al.* analyzed how young people's thinking styles are related to their awareness and skills in evaluating the credibility of online information [16]. In addition, researchers in IR community found that people with different thinking styles behave differently in the Web search process. Wu *et al.* reported that people with high NFC are likely to spend more time on a search task, click deeper in search results, and more paginations [27].

In this research, we focused on the following four thinking styles:

- Need for Cognition (NFC) [3]: how much a person enjoys effortful cognitive activities.
- Flexible Thinking (FT) [1]: how a person is willing to change their beliefs in the face of contradictory evidence.

- Faith in Intuition (FII) [8]: how a person relies on an intuitive decision making.
- General Trust (GT) [28]: a person's honesty and trustworthiness in general.

Note that the above three thinking styles, NFC, FT, and FII were adapted from the study of Metzger *et al.* [16]. Given the characteristics of these thinking styles, we propose the following hypotheses:

> **H1-2** Need for Cognition is positively correlated to the verification attitude.
>
> **H1-3** Flexible Thinking is positively correlated to the verification attitude.
>
> **H1-4** Faith in Intuition is negatively correlated to the verification attitude.
>
> **H1-5** General Trust is negatively correlated to the verification attitude.

3.2.3 Educational background. Many studies have reported that the experiences of learning activities at the university level develop students' information literacy skills and critical thinking abilities [15], based on which we propose the following hypothesis:

> **H1-6** People with high educational background are likely to have stronger verification attitudes.

3.2.4 Search expertise. Search expertise would play an important role in determining whether people have strong verification attitudes. Given that search experts know more about a Web search engine than the other people, they can use several verification strategies when using a Web search engine. For example, the search experts can easily use a search tool to filter Web pages that have been recently updated to see up-to-date information. We thus propose the following hypothesis:

> **H1-7** Search experts are likely to have stronger verification attitudes.

3.3 Behavioral Differences in Web Search

To address the research question **RQ2**, we discuss the possible differences in people's Web search behaviors. While some verification strategies can be evaluated by query log analysis of a search engine, others are quite difficult to verify, and a laboratory study might be more suitable. For example, verifying whether a person tries to issue multiple queries in a session can be addressed if their query logs are available, verifying whether a person tries to be aware the goal of the author of a Web page would be difficult based only on query log analysis. The present study focuses on relatively shallow behavioral signals, for which a large-scale query log analysis can be applied. The rest of this subsection discusses such behavioral signals and our expectations of how they differ according to people's verification attitudes.

3.3.1 Time. In general, implementing the verification strategies, as described in Section 3.1, requires a person to spend more time on the search process. For example, browsing multiple Web pages to scrutinize information and issuing multiple queries to collect information from different perspectives can be typical strategies. We thus propose the following two hypotheses:

H2-1 People ranked high on verification attitude spend more time on a search session.

H2-2 People ranked high on verification attitude issue more queries in a search session.

3.3.2 Clicks. Similar to the discussion in Section 3.3.1, a person ranked high on verification attitude may try to check a greater number of pages than the other people; thus we first propose:

H2-3 People ranked high on verification attitude click a greater number of search results in a session.

In addition to the number of clicks in a session, we hypothesize that people ranked high on verification attitude are tolerant to the position bias. Such people may click not only the top-ranked search results but also lower-ranked results to collect multiple evidences for verifying information. We thus, propose the hypothesis:

H2-4 People ranked high on verification attitude click lower-ranked search results.

3.3.3 Queries. When people encounter uncertain or incredible information on the Web, people ranked high on verification attitude may seek out the evidences to scrutinize the information. In such a situation, people may issue a query that can retrieve evidences. We assume that such a query contains the terms that are likely to be used to verify or scrutinize information. Such terms include "evidence," "data," and "truth," *etc.* We refer to these terms as *verification terms* (See Section 5.5 and Table 8 for the details). We expect that people ranked high on verification attitude are likely to use the verification terms in their queries, and thus, propose the following hypothesis:

H2-5 People ranked high on verification attitude are more likely to issue queries that contain the verification terms.

Hypotheses **H1-1** to **H1-7** will be investigated by conducting an online survey (See Section 4). Hypotheses **H2-1** to **H2-5** will be investigated by analyzing the query logs of the participants of the online questionnaire (See Section 5).

4 ONLINE QUESTIONNAIRE

To answer hypotheses **H1-1** to **H1-7**, we administered an online questionnaire. We first explain the participants and the procedure of the questionnaire. We then introduce the questionnaire items to measure the factors described in Section 3.2. Finally, we report the results and discuss whether the hypotheses are supported.

4.1 Participants and Procedure

The participants of the online questionnaire were recruited via Yahoo Japan Crowdsourcing[1], a crowdsourcing platform in Japan. We first informed the participants that the results of the questionnaire would be analyzed by relating them with their search behaviors on the Web search engine operated by Yahoo Japan. Only those who agreed to this condition were administered the questionnaire described in Section 4.2.

The questionnaire was administered in November 2017. The questionnaire took approximately 15 minutes to complete. Each participant received approximately 0.5 USD for completing this questionnaire. In total, 3,621 people were administered the questionnaire.

[1]https://crowdsourcing.yahoo.co.jp/

Table 1: Demographics of participants.

Gender	n	Age	n	Educational Background	n
male	835	10s	13	university-educated	909
female	637	20s	124	non-university-educated	549
N/A	19	30s	422	N/A	33
		40s	612		
		50s or older	301		
		N/A	19		

We used only the responses of the 1,491 participants, who used our Web search engine on a daily basis (See Section 5.1). The demographics of the 1,491 participants are summarized in Table 1.

4.2 Questionnaires

We prepared the following questionnaire items, as listed in Table 2:

- Credulity toward Web information and search results
- Attitude toward using verification strategies in Web search
- Thinking styles
- Demographics and search expertise

4.2.1 Credulity toward Web information and search results. Thirteen items were prepared to measure the tendency of people to believe the information encountered on the Web. The first 10 items were adapted from the studies in [16, 17] to measure how people are likely to believe information on the Web, including the following:

- How much of the information available on the Web do you think is believable? (from 1:"none of it" to 5: "all of it.")
- How likely are you to believe various types of information, such as news, entertainment, health, facts, education, shopping, recreation, and information from online friends you find on the Web? (from 1:"not at all likely to believe it" to 5:"very likely to believe it.")

The other three items were prepared to measure how a person is likely to believe the search results returned by a Web search engine, including the following:

- How likely do you believe the information contained in the Web pages returned by a Web search engine? (from 1:"not at all likely to believe it" to 5:"very likely to believe it.")
- How likely are you to believe the high-ranked (low-ranked) search results returned by a Web search engine? (from 1:"not at all likely to believe it" to 5:"very likely to believe it.")

We regarded that the people who responded with high scores on these 13 items have high tendency to believe the information on the Web.

After we obtained the responses of the 1,491 participants, Cronbach's α coefficient [5] was measured for the above 13 items. The Cronbach's α coefficient measures the internal consistency of the responses among a set of items, and it is used to validate the reliability of the items. The Cronbach's α coefficient of the 13 items was 0.84, which means these items are reliable for measuring the same characteristic.

4.2.2 Attitude toward using verification strategies in Web search. Fifteen items were prepared to measure a person's verification attitude. To this end, we asked the participants whether they try to

Table 2: Questionnaire items used in online questionnaire. Items marked with *(r)* are reverse coded.

Category	Item	M	S.D.
Credulity toward Web information and search results	How much of the information available on the Web do you think is believable?	3.24	0.73
	How believable do you feel is the information on the Web?	3.39	0.62
	How likely are you to believe news you find on the Web?	3.67	0.71
	How likely are you to believe the information about entertainment such as sports, celebrity, or hobbies you find on the Web?	3.20	0.87
	How likely are you to believe the information about health or disease you find on the Web?	3.01	0.10
	How likely are you to believe the information about factoids you find on the Web?	4.16	0.81
	How likely are you to believe the information about education you find on the Web?	3.32	0.85
	How likely are you to believe the information about shopping you find on the Web?	3.75	0.81
	How likely are you to believe the information about recreation such as travel you find on the Web?	3.92	0.75
	How likely are you to believe the information shared by other online friends?	2.86	1.01
	How likely are you to believe the information contained in the web pages returned by a Web search engine?	3.49	0.81
	How likely are you to believe the high-ranked (approx. 1st-5th) search results returned by a Web search engine?	3.53	0.82
	How likely are you to believe the lower-ranked (approx. 5th-10th) search results returned by a Web search engine?	3.03	0.91
Attitude toward using verification strategies	I try to check whether the information on the page is up-to-date.	3.14	0.87
	I try to check whether the information on the page is complete and comprehensive.	2.65	0.86
	I try evaluating whether the views represented on the page are facts or opinions.	3.22	0.96
	I try to consider the author's goals/objectives in posting information on the page.	2.95	0.99
	I try to identify the author of the page.	2.68	1.05
	I try to look for a stamp of approval or recommendation from third parties on the Web page.	2.65	0.93
	I try to check whether the author's contact information is provided.	2.47	0.94
	I try to verify the author's qualifications or credentials on the page.	2.40	0.94
	I try to click multiple Web pages when using a search engine.	3.56	0.91
	I try to browse both lower- and higher-ranked search results	2.94	0.93
	I try to click search results in which I can easily identify the author of the page, such as an official site of a company.	3.27	0.98
	I try to check the domains of search results before clicking them.	2.24	1.01
	I try to issue multiple queries in general.	3.54	0.89
	I try to use a search tool to filter recently updated Web pages .	3.20	0.96
	I try to spend as much time as possible on Web searches.	3.20	0.90
Thinking Styles			
Need for Cognition (NFC)	Thinking is not my idea of fun. *(r)*	4.04	0.93
	I try to anticipate and avoid situations in which I might be required to think in depth about something. *(r)*	3.25	1.08
	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. *(r)*	3.51	1.03
	I would prefer complex to simple problems.	2.85	1.01
	I find little satisfaction in deliberating hard and for long hours. *(r)*	3.47	0.94
Faith in Intuition (FII)	I trust my initial feelings about people.	3.01	0.83
	I believe in trusting my hunches.	3.02	1.00
	My initial impressions of people are almost always right.	3.13	0.80
	When it comes to trusting people, I can usually rely on my "gut feelings."	3.27	0.92
	I can usually feel when a person is right or wrong even if I cannot explain how I know it.	3.20	0.88
General Trust (GT)	Most people are trustworthy.	2.74	0.99
	Most people will respond in kind when they are trusted by others.	3.31	0.95
	Most people are trustful of others.	2.74	0.95
	Most people are basically honest.	2.80	1.02
	I am trustful.	3.16	1.03
	Most people are basically good and kind.	2.92	1.01

adopt any strategies for verifying information during their Web search process. The first eight items were adapted from the study by Metzger *et al.* [17] to measure the attitude toward using the verification strategies when browsing Web pages, including the following:

- I try to check whether the information on the page is up-to-date.
- I try to identify the author of the page.

In addition to the above items, we also prepared seven items to measure the attitude toward using verification strategies when performing Web search using a search engine. Such items include:

- I try to browse both lower and higher ranked search results.
- I try to spend as much time as possible on Web searches.

All 15 items were answered on a five-point Likert scale (from 1: "never" to 5: "almost always.").

We regarded the people who responded with high scores for these 15 items have strong verification attitudes. The Cronbach's α coefficient of the 15 items was 0.89.

4.2.3 Thinking styles. Need for Cognition was measured based on five items adapted from [8]. Faith in Intuition (FII), too, was measured based on five items adapted from [8]. Flexible Thinking (FT) was measured based on 10 items adapted from [21]. General Trust (GT) was measured based on six items adapted from [28]. All items related to thinking styles were answered on a five-point Likert scale (from 1: "strongly disagree" to 5: "strongly agree"). Note that the order of presentation of the items related to the thinking styles were randomized to avoid the order effect.

The Cronbach's α coefficients of the above thinking styles were 0.74 (NFC), 0.70 (FII), 0.20 (FT), and 0.86 (GT), respectively. Because the Cronbach's α coefficient of FT was quite low, indicating the responses to the items were inconsistent and thus not reliable, we excluded FT from our analysis. One possible reason why we obtained the low α for FT is that the nuances of the items were changed when we translated the items in [21] into Japanese.

4.2.4 Demographics and search expertise. Basic demographics including gender, age, occupation, and educational background were

Table 3: Correlations between factors ($^*p < 0.01$, $^{**}p < 0.05$).

	Attitude	Credulity	NFC	FII	GT
Attitude	–	−0.08**	0.28**	0.00	0.03
Credulity		–	−0.07*	0.11**	0.26**
NFC			–	−0.09**	−0.00
FII				–	0.24**

asked in the questionnaire. As for the educational background, if a participant answered that they were university/college graduates, they were considered "university-educated" ($n = 909$); otherwise, they were considered "non-university-educated" ($n = 549$). Note that the participants had the option to refrain from answering the demographic questions.

In addition, we asked a question pertaining to people's search expertise in the questionnaire. Following the study of White and Morris [26], we regarded the participants who had experienced in using advanced query operators as search experts:

- Do you have any experience of using options such as NOT or double-quoted phrases when using the Web search engine? ("yes" or "no")

The participants who answered "yes" to this question were treated as "search experts" ($n = 457$), and the others were treated as "non-experts" ($n = 1,034$) in our analysis.

4.3 Results

Table 2 lists the means and standard deviations of the results for each questionnaire item. To verify hypotheses **H1-1** to **H1-5**, we first calculated the verification attitude score for each participant by averaging their scores of the items on the verification attitude. Similarity, we calculated the scores for the other factors for each participant by averaging the scores of the items relating to the factors. We then computed the correlations between the verification attitudes and the other factors. Table 3 lists the Pearson's correlation coefficients among factors. From the results, we observed a positive correlation between verification attitudes and NFC ($r = 0.28$). Because NFC reflects how a person enjoys complex activities with high cognitive load, people ranked higher on NFC are more willing to use strategies for verifying information in Web search. Contrary to our expectation, the correlation between the verification attitude and credulity is quite weak ($r = −0.08$). Similarly, we did not observe any significant correlation between the verification attitudes and FII ($r = 0.00$) and GT ($r = 0.03$).

In addition to verification attitude, we observed a positive correlation between credulity and GT ($r = 0.26$). This results is reasonable because people who generally trust others would be less likely to question about the credibility of the information on the Web.

To verify hypotheses **H1-6** and **H1-7**, we ordered the results by educational background and search expertise. Table 4 lists the results of verification attitudes, credulity, and thinking styles by educational background and search expertise. As for educational background, a significant difference in verification attitudes between university-educated and non-university-educated participants was identified by conducting Welch's t-test ($t = 3.30$, $p < 0.01$). Moreover, a statistically significant difference in verification

attitudes between search experts and non-experts was found ($t = 11.84$, $p < 0.01$).

4.4 Summary of Questionnaire Analysis

In summary, hypothesis **H1-2** was supported by the results of our analysis. The hypotheses pertaining to educational background (**H1-6**) and search expertise (**H1-7**) were also supported by the results of our analysis. By contrast, we found no evidence to support hypotheses **H1-1**, **H1-3**, **H1-4**, and **H1-5**.

5 QUERY LOG ANALYSIS

To verify hypotheses **H2-1** to **H2-5**, we analyzed the query logs of the participants who were administered the online questionnaire in Section 4. We first explain the query logs and the two datasets we prepared herein. We then report the analysis results.

5.1 Data Collection

We used the query logs recorded by the Yahoo Japan Web search engine, one of the most popular Web search engine in Japan. We used the one-year and one-month query logs recorded from October 1, 2016 to October 31, 2017. The logs contained only the search queries issued on desktop devices and not those issued on mobile and tablet devices. Moreover, the query logs used herein did not contain the records of abandoned queries that were not followed by any click.

To collect the query logs used in the analysis, we first extracted the query logs of the 3,621 participants who were administered the online questionnaire. We then split their logs into sessions. Here, a session was defined based on a 30-minutes inactivity [24]. To remove the logs of users who did not use the Web search engine, we only selected the logs of the participants that contained 10 or more sessions. As a result, we collected the query logs of 1,491 users, which amounted to 560,174 sessions. We call this dataset NORMAL hereinafter. Note that the results of the questionnaire of these 1,491 users were reported in Section 4.

One problem of NORMAL is that it contains sessions of *navigational* queries [2], or simple fact-finding search tasks, for which people need not use verification strategies in the search process in general. To mitigate this problem, we also prepared a dataset containing only longer search sessions. More precisely, from the NORMAL dataset we extracted sessions in which a participant issued three or more queries. From the remaining logs, we eliminated those logs that contained fewer than five sessions. As a result, we extracted the query logs of 1,122 users, which amounted to 154,701 long sessions.[2] We call this dataset LONG hereinafter.

5.2 Categorizing Participants

To verify hypotheses **H2-1** to **H2-5**, we first classified the participants according to the verification attitude scores obtained from the online questionnaire. The participants whose verification attitude scores were in the top 10 percentile were referred to as *high* verification attitude participants ($n = 137$ for NORMAL, $n = 106$

---

[2]Due to the space limitation the questionnaire results of the 1,122 are not reported. We note that the questionnaire results of the 1,122 users showed the similar trend as those of 1,491 users in terms of the significant differences between factors.

Table 4: Results of verification attitudes, credulity, and thinking styles by educational background and search expertise ($^*p < 0.05,^{**}p < 0.01$).

		Attitude		Credulity		NFC		FII		GT	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Educational background	university-educated ($n = 909$)	2.98**	0.60	3.40**	0.49	3.46**	0.70	3.10*	0.60	2.97	0.76
	non-university-educated ($n = 549$)	2.88	0.57	3.49	0.45	3.37	0.70	3.18	0.59	2.91	0.77
Search expertise	search experts ($n = 457$)	3.20**	0.56	3.37**	0.47	3.55**	0.72	3.09	0.66	2.91	0.75
	non-experts ($n = 1,034$)	2.83	0.57	3.45	0.48	3.37	0.69	3.14	0.57	2.96	0.77

for LONG) by following the methodology of Wu *et al.* [27]. Similarly, the participants whose verification attitude scores were between the top 11 to 50 percentile were referred to as *medium* verification attitude participants ($n = 561$ for NORMAL, $n = 454$ for LONG), and the others were referred to as *low* verification attitude participants ($n = 793$ for NORMAL, $n = 562$ for LONG). We also categorized the participants based on the other factors such as credulity, NFC, FII and GT, by using the same procedure.

5.3 Basic Statistics

To verify the hypotheses, we first examined the basic statistics of search sessions such as the number of queries, dwell time, and number of clicks, by verification attitude. Tables 5 and 6 summarize these results for different people in terms of their verification attitudes in the NORMAL and the LONG datasets, respectively.

In case of the LONG dataset, there was a statistically significant difference in the number of queries between groups as determined by one-way ANOVA ($F(2, 1119) = 3.14, p < 0.05$). A post-hoc Tukey's HSD test showed that the difference in the number of queries between the *medium* and *low* verification attitude participants is significant ($p < 0.05$). However, there were no other statistically significant differences between user group means as determined by one-way ANOVA, for both the NORMAL and the LONG datasets. While we expected that the participants ranked high on verification attitude were likely to issue more queries (**H2-1**), spend more time (**H2-2**), and click on more results (**H2-3**), no evidence supporting these hypotheses was found from the analysis.

We also investigated whether there exist differences in other factors such as credulity, NFC, FII, GT, educational-background, and search expertise, but no significant difference in those behavioral signals was found for each of the factors.

5.4 Click Positions

To verify hypothesis **H2-4**, we investigated the positions (i.e., *rank*) of the clicked search results. Table 7 shows the deepest position of the clicked search results in a session for the participants with different verification attitudes. There were statistically significant differences between groups, as determined by one-way ANOVA for both the NORMAL ($F(2, 1488) = 3.78, p < 0.05$) and the LONG ($F(2, 1119) = 3.90, p < 0.05$) datasets. A post-hoc Tukey's HSD test showed that the differences between *high* and *medium* verification attitude participants were significant for both the NORMAL ($p < 0.05$) and the LONG ($p < 0.05$) datasets. In addition, we observed that in terms of educational background and search expertise, no statistically significant differences were determined

Table 5: Number of queries, dwell time, and number of clicks per session by attitudes for NORMAL dataset. No significant differences were identified by one-way ANOVA ($p < 0.05$).

	Attitude					
	high		medium		low	
	M	S.D.	M	S.D.	M	S.D.
number of queries	2.05	1.17	2.10	1.52	2.02	0.96
dwell time (sec)	1,668	1,877	1,884	2,593	2,030	2,624
number of clicks	3.23	1.90	3.17	2.08	3.14	1.52

Table 6: Number of queries, dwell time, and number of clicks per session by verification attitudes for LONG dataset. Difference in number of queries between medium and low participants was significant ($p < 0.05$).

	Attitude					
	high		medium		low	
	M	S.D.	M	S.D.	M	S.D.
number of queries	4.80	1.56	4.98	2.27	4.69	1.55
dwell time (sec)	2,994	2,722	3,304	3,165	3,383	3,094
number of clicks	7.93	2.81	7.86	3.54	7.55	2.52

Table 7: Deepest position of clicked search results per session by verification attitudes in NORMAL and LONG datasets. Differences between high and medium participants are significant ($p < 0.05$) for both NORMAL and LONG datasets.

	Attitude					
	high		medium		low	
	M	S.D.	M	S.D.	M	S.D.
NORMAL	4.94	5.00	4.15	2.79	4.30	2.68
LONG	9.58	8.07	7.94	5.02	8.33	5.36

in Welch's t-test for both the NORMAL (educational background: $t = 1.65, p = 0.29$, search expertise: $t = 0.01, p = 1.00$) and the LONG (educational background: $t = 1.16, p = 0.24$, search expertise: $t = 0.11, p = 0.91$) datasets.

We then investigated the position distribution of the clicked search results. Figure 1 shows the position distributions of the clicked search results by participants of different verification attitudes for the NORMAL and the LONG datasets. From the figure, we can see that the participants ranked high on verification attitude were *less* likely to click the top-ranked search result. Given the similar numbers of clicks per session among the different verification attitude groups, as in Tables 5 and 6, a possible reason for the difference in
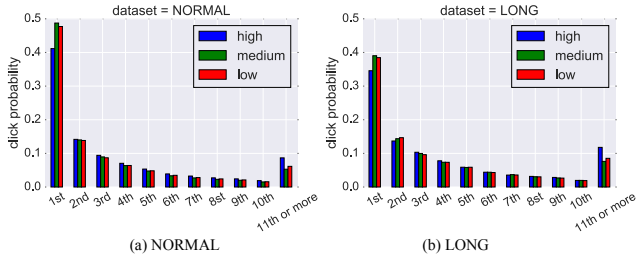
(a) NORMAL (b) LONG

Figure 1: Position distributions of clicked search results by verification attitude for (a) NORMAL and (b) LONG datasets.
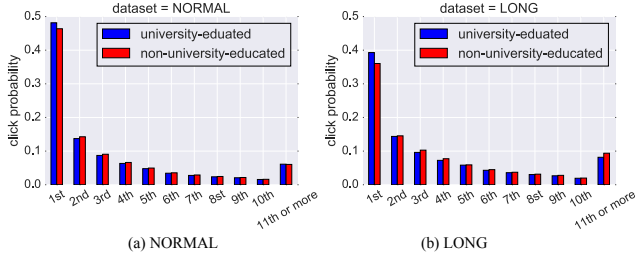

(a) NORMAL (b) LONG

Figure 2: Position distributions of clicked search results by participants in different educational backgrounds for (a) NORMAL and (b) LONG datasets.

(a) NORMAL (b) LONG

Figure 3: Position distributions of clicked search results by participants in different search expertise for (a) NORMAL and (b) LONG datasets.

(a) NORMAL (b) LONG

Figure 4: Position distributions of clicked search results by people in different credulity for (a) NORMAL and (b) LONG datasets.

the position distributions is that the people ranked high on verification attitude are less biased in terms of the position of the search results returned by a search engine and thus click lower-ranked search results instead of clicking the top-ranked ones.

In addition, we investigated participants' position distributions based on their educational backgrounds and search expertise, because these two factors are positively correlated to the verification attitudes, as found in Section 4.3. Figures 2 and 3 show the position distributions for the participants in different educational backgrounds and search expertise, respectively. Regarding educational background, contrary to our expectation, we cannot find a similar trend as in case of the verification attitude groups. For search expertise, we can see a similar tendency with that in case of verification attitudes, although the difference appears to be smaller.

Apart from the verification attitudes, we investigated the position distributions of the participants with different credulity because people's position bias may strongly be affected by the extent to which believe the search results returned by a search engine. Figure 4 shows the position distributions of the participants with different credulity. Figure 4 shows that the participants high on credulity appear to rely strongly on the top-ranked search results than participants low on credulity, possibly because the position bias among such participants was stronger than that among the other participants.

5.5 Queries for Verification

To verify hypothesis **H2-5**, we first prepared the verification terms. To this end, we first prepared the 17 terms used in the Yamamoto *et al.*'s study [31]. They used a crowd sourcing to collect terms in which people who have high critical thinking skills and dispositions are likely to use in their query. Table 8 lists the verification terms considered in the present study. Seventeen terms were
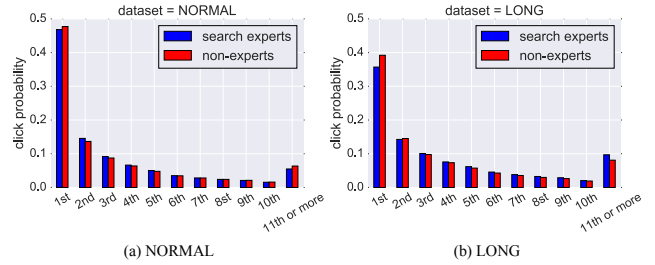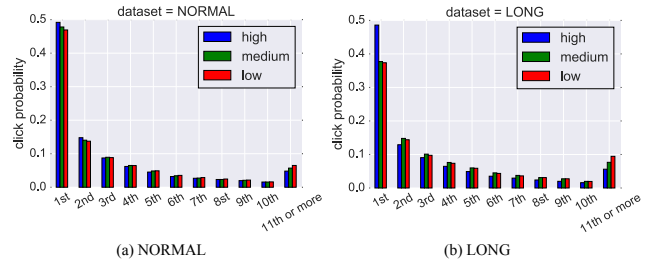
adapted from their study [31]. In addition, we also manually added five terms and used these 22 terms as the verification terms.

We then computed the probability that a participant formulated a query containing one of the verification terms in a session. Because the probability is very small, to emphasize the relative difference from the average users, we normalized the probability against the average probability of all users in the dataset. We refer to this probability as relative ratio. Figure 5 shows the relative ratios of the participants in different verification attitudes for the NORMAL and LONG datasets. The figure shows that the participants ranked high on verification attitude were more likely to use the verification terms in their query. The data suggest that the participants ranked high on verification attitude were likely to use the verification terms approximately 1.3 times the average, while the participants ranked low on verification attitude were likely to use the verification terms 0.9 times the average. A Chi-square test revealed that the percentage of using the verification terms in their query significantly differed by their verification attitudes for both the NORMAL ($\chi^2 = 86.03$, $p < 0.01$) and the LONG ($\chi^2 = 83.73$, $p < 0.01$) datasets. Figures 6 and 7 show the relative ratios of the participants in different educational backgrounds and search expertise, respectively. Regarding educational background, we observe that the university-educated participants seemed less likely to use the verification terms (only significant for the NORMAL dataset ($p < 0.01$)). Regarding search expertise, we can see that the search experts were more likely to use the verification terms (significant for both the NORMAL ($p < 0.01$) and the LONG ($p < 0.01$) datasets), while the difference between search experts and non-experts appear to be smaller than that between the participants ranked high on verification attitude and those ranked low on verification attitude.
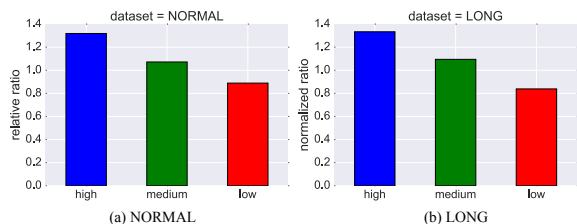
Figure 5: Relative ratio of use of verification terms in a session by verification attitude for (a) NORMAL and (b) LONG datasets.
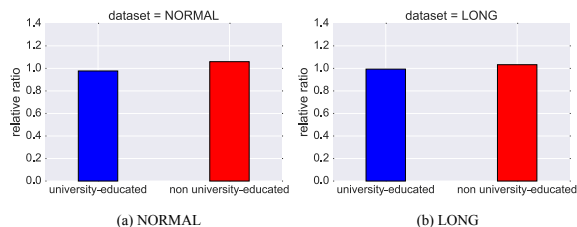

Figure 6: Relative ratio of use of verification terms in a session by participants in different educational backgrounds for (a) NORMAL and (b) LONG datasets.
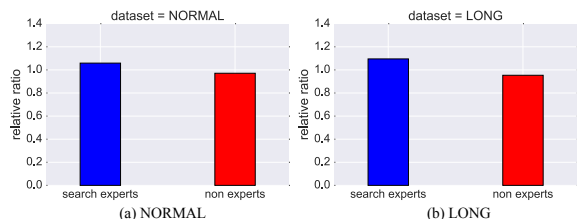

Figure 7: Relative ratio of use of verification terms in a session by participants in different search expertise for (a) NORMAL and (b) LONG datasets.

Table 8: Verification terms used in our analysis. Underlined terms were adapted from study [31].

Verification terms
principle, evidence, mechanism, process, proof, survey, research, validation, pursuit, comparison, stats, analysis, difference, reputation, evidence, actual proof, data, lie, effect, reason, what reason, truth

5.6 Summary of Query Log Analysis

In summary, we found no evidence supporting hypotheses **H2-1** to **H2-3** from the results of our query log analysis. As for hypothesis **H2-4**, we found that the participants ranked high on verification attitude clicked on deeper search results than those with intermediate level of verification attitude. The position distribution for the participants from different verification attitude groups also supported this hypothesis. Finally, hypothesis **H2-5** was supported from the results of our query log analysis.

6 DISCUSSIONS

Our analysis of the questionnaire and query logs revealed how people's attitudes toward using verification strategies in Web search are related to several factors, as well as the relationship between verification attitudes and people's daily search behaviors. We first discuss the implications of the finding of our study, and then, we mention the limitations of this study.

6.1 Implications

The results of the query log analysis showed that the participants high on verification attitude clicked lower-ranked search results (Section 5.4), and were more likely to issue queries containing the verification terms (Section 5.5). One interesting finding of our study is that such behavioral differences were not identified from the differences in the participants' educational backgrounds, although the university-educated participants were likely to have higher verification attitudes (Section 4.3).

One possible explanation for this is that the *skills* required when people try to obtain accurate and credible information via Web search and those required in general information evaluation would be different. For example, the tolerance of the position bias, use of the advanced search tools to obtain up-to-date information can be considered as skills in verifying information unique to Web search. Therefore, the search behaviors of the university-educated participants, who are expected to have good information literacy skills and critical thinking skills [15], did not show the behavioral differences as found in the participants in different verification attitudes. In such a case, an educational program specially designed for Web search would be important, even for highly educated people, to help them obtain accurate information via Web search.

Another possible explanation for this is that the verification *attitudes* of the university-educated participants were insufficient to expend efforts on using verification strategies in Web search. Although the university-educated participants are likely to have higher verification attitudes, the difference between university-educated and non-university-educated participants are small, compared to that between search experts and non-experts (Section 4.3). Therefore, while the university-educated participants might have sufficient skills in verifying information, their attitudes might be insufficient to expend efforts on verification strategies. In such a case, we need a search system that encourage the participants' verification *attitudes*, such as proposed in [6, 29, 31].

6.2 Limitations

This study has several limitations. First, many factors affect people's search behavior, and such effects were not fully separated from our query logs. Especially, search task types and search expertise strongly affect people's behaviors in Web search.

As for search task types, although we prepared a dataset focused on longer search sessions, this dataset still contained various types of search sessions. Some search sessions, such as intrinsically diverse search tasks [19] or exploring sessions [9], required a searcher to expense efforts on verification strategies, while others did not require the searcher to use the verification strategies but required the searcher to be skilled in formulating a successful query, such as struggling search [9]. To focus only on the search sessions in which the searchers are required to carefully use the Web search engine and critically evaluate the information, we need to further choose such sessions, for example, by applying the algorithm proposed by Raman *et al.* [19] to the query logs to obtain the sessions in which people seek for the diverse information.

As for search expertise, usually, search experts are expected to formulate an effective query that retrieves the relevant Web page at the higher position, which means that the search experts are expected to issue less queries, spend less time, and click less search results in a session [26], all of which contrary to the hypotheses we prepared for the participants with high verification attitudes (**H2-1** to **H2-3**). Our query log analysis does not suffice to distinguish the effects between search expertise and verification attitudes. Also, there may exist an interaction effect between search expertise and verification attitudes on search behavior, in which the current analysis does not take into consideration. To better understand the relationship between people's verification attitudes and their behaviors, we need to eliminate the effects of search expertise from our analysis.

Second, people who use good verification strategies in their Web search process are not necessarily good at obtaining *accurate* information via Web search. Metzger *et al.* found that young students who have greater exposure to credibility evaluation training are more likely to believe hoax Web pages [16]. Given this fact, it is possible that people ranked high on verification attitude and use of verification strategies are more likely to believe hoax Web pages that are seemingly credible via Web search. To better understand people's behaviors and the skills in seeing through inaccurate information on the Web, we need to conduct a laboratory study, as well as a questionnaire-based survey and query log analysis.

7 CONCLUSION

In this study we investigated how people carefully search for the Web to obtain credible and accurate information. Our online questionnaire revealed that the people's attitudes toward using verification strategies in Web search differ based on several factors such as Need for Cognition, educational background, and search expertise. Furthermore, we found that the people's search behaviors in terms of the positions of the clicked search results and the likelihood of issuing queries for verifying or scrutinizing information differ based on their verification attitudes.

The findings of this study imply that people's search behaviors are different according to their verification attitudes. As we discussed in Sections 3.3 and 6.2, even a combination of questionnaire and query log analysis is insufficient to fully understand people's verification behaviors in Web search, for example, how do they seek information to find evidence for validating information or their awareness of the author's goals in a given Web page. In the future, we plan to incorporate a laboratory study into our analysis.

REFERENCES

[1] Jonathan Baron. 2000. *Thinking and deciding.* Cambridge University Press.
[2] Andrei Broder. 2002. A taxonomy of web search. 36, 2 (2002), 3–10.
[3] John T Cacioppo, Richard E Petty, Jeffrey A Feinstein, and W Blair G Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Review* 119, 2 (1996), 197–253.
[4] Michael J Cole, Xiangmin Zhang, Chang Liu, Nicholas J Belkin, and Jacek Gwizdka. 2011. Knowledge effects on document selection in search results pages. In *SIGIR*. 1219–1220.
[5] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 3 (1951), 297–334.
[6] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. 2010. Highlighting disputed claims on the web. In *WWW*. 341–350.
[7] Robert H Ennis. 1987. A taxonomy of critical thinking dispositions and abilities. (1987).
[8] Seymour Epstein, Rosemary Pacini, Veronika Denes-Raj, and Harriet Heier. 1996. Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of personality and social psychology* 71, 2 (1996), 390.
[9] Ahmed Hassan, Ryen W White, Susan T Dumais, and Yi-Min Wang. 2014. Struggling or exploring?: disambiguating long search sessions. In *WSDM*. 53–62.
[10] Helene A Hembrooke, Laura A Granka, Geraldine K Gay, and Elizabeth D Liddy. 2005. The effects of expertise and feedback on search term selection and subsequent learning. *Journal of the Association for Information Science and Technology* 56, 8 (2005), 861–871.
[11] Christoph Hölscher and Gerhard Strube. 2000. Web search behavior of Internet experts and newbies. *Computer networks* 33, 1-6 (2000), 337–346.
[12] Patricia Iannuzzi. 2000. Information literacy competency standards for higher education. *Community & Junior College Libraries* 9, 4 (2000), 63–67.
[13] Adobe Inc. 2015. The State of Content: Rules of Engagement for 2016. https://www.adobe.com/content/dam/acom/en/news-room/pdfs/201512/state-of-content-report.pdf
[14] Dae-Young Kim, Xinran Y Lehto, and Alastair M Morrison. 2007. Gender differences in online travel information search: Implications for marketing communications on the internet. *Tourism management* 28, 2 (2007), 423–433.
[15] Marc Meola. 2004. Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *portal: Libraries and the Academy* 4, 3 (2004), 331–344.
[16] Miriam J Metzger, Andrew J Flanagin, Alex Markov, Rebekah Grossman, and Monica Bulger. 2015. Believing the unbelievable: understanding young people's information literacy beliefs and practices in the United States. *Journal of Children and Media* 9, 3 (2015), 325–348.
[17] Miriam J Metzger, Andrew J Flanagin, and Lara Zwarun. 2003. College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education* 41, 3 (2003), 271–290.
[18] Satoshi Nakamura, Shinji Konishi, Adam Jatowt, Hiroaki Ohshima, Hiroyuki Kondo, Taro Tezuka, Satoshi Oyama, and Katsumi Tanaka. 2007. Trustworthiness analysis of web search results. In *ECDL*. 38–49.
[19] Karthik Raman, Paul N Bennett, and Kevyn Collins-Thompson. 2013. Toward whole-session relevance: exploring intrinsic diversity in web search. In *SIGIR*. 463–472.
[20] Elizabeth Sillence, Pam Briggs, Lesley Fishwick, and Peter Harris. 2004. Trust and mistrust of online health sites. In *CHI*. 663–670.
[21] Keith E Stanovich and Richard F West. 1997. Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology* 89, 2 (1997), 342.
[22] Ingmar Weber and Alejandro Jaimes. 2011. Who uses web search for what: and how. In *WSDM*. 15–24.
[23] Ryen W White. 2013. Beliefs and biases in web search. In *SIGIR*. 3–12.
[24] Ryen W White and Steven M Drucker. 2007. Investigating behavioral variability in web search. In *WWW*. 21–30.
[25] Ryen W White, Susan T Dumais, and Jaime Teevan. 2009. Characterizing the influence of domain expertise on web search behavior. In *WSDM*. 132–141.
[26] Ryen W White and Dan Morris. 2007. Investigating the querying and browsing behavior of advanced search engine users. In *SIGIR*. 255–262.
[27] Wan-Ching Wu, Diane Kelly, and Avneesh Sud. 2014. Using information scent and need for cognition to understand online search behavior. In *SIGIR*. 557–566.
[28] Toshio Yamagishi and Midori Yamagishi. 1994. Trust and commitment in the United States and Japan. *Motivation and emotion* 18, 2 (1994), 129–166.
[29] Yusuke Yamamoto and Satoshi Shimada. 2016. Can disputed topic suggestion enhance user consideration of information credibility in web search?. In *HT*. 169–177.
[30] Yusuke Yamamoto and Katsumi Tanaka. 2011. Enhancing credibility judgment of web search results. In *CHI*. 1235–1244.
[31] Yusuke Yamamoto and Takehiro Yamamoto. 2018. Query priming for promoting critical thinking in web search. In *CHIIR*. 12–21.
[32] Yusuke Yamamoto, Takehiro Yamamoto, Hiroaki Ohshima, and Hiroshi Kawakami. 2018. Web Access Literacy Scale to Evaluate How Critically Users Can Browse and Search for Web Information. In *WebSci*. 97–106.
[33] Yisong Yue, Rajan Patel, and Hein Roehrig. 2010. Beyond position bias: Examining result attractiveness as a source of presentation bias in clickthrough data. In *WWW*. 1011–1018.
[34] Mingming Zhou. 2014. Gender difference in web search perceptions and behavior: Does it vary by task performance? *Computers & Education* 78 (2014), 174–184.