Disputed sentence suggestion towards credibility-oriented Web search

Yusuke Yamamoto

Kyoto Univeristy, Yoshida-honmachi, Kyoto, Japan
yamamoto@dl.kuis.kyoto-u.ac.jp

Abstract. We propose a novel type of query suggestion to support credibility-oriented Web search. When users issue queries to search for Web pages, our system collects disputed sentences about queries from the Web. Then, the system measures how typical and relevant each of the collected disputed sentences are to the given queries. Finally, the system suggests some of the most typical and relevant disputed sentences to the users. Conventional query suggestion techniques focus only on making it easy for users to search for Web pages matching their intent. Therefore, when users search for Web pages to check the credibility of specific opinions or statements, queries suggested by conventional techniques are not always useful in searching for evidence for credibility judgments. In addition, if users are not careful about the credibility of information in the Web search process, it is difficult to be aware of the existence of suspicious Web information through conventional query suggestions. Our disputed sentence suggestion enhances users' awareness of suspicious statements so that they can search for Web pages with careful attention to them.

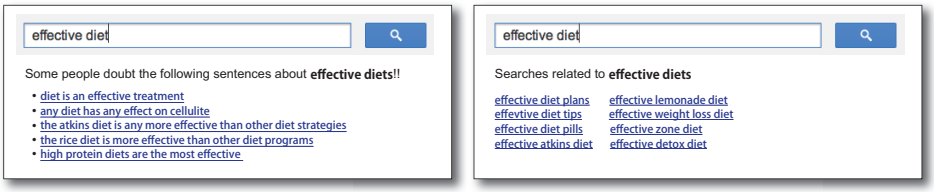
1 Introduction

Nowadays, the Web is a huge information resource. People can easily and freely obtain information through Web search engines and specific Web sites. However, the problem of Web information credibility is emerging [1, 2].

The quality control of Web information is generally insufficient due to low publishing barriers. As a result, there is a large amount of incorrect and unreliable information on the Web that can have detrimental effects on users. For instance, Sillence et al. reported that there are more than 20,000 medical Web sites on the Web, but over half of them have not been checked by medical experts [1]. On the other hand, Nakaura et al. reported that a lot of Web users trust Web information to some extent [2]. In particular, for Web search engines, they also reported that more than 50% of search engine users perceive the search results to be somewhat credible.

Although Web information credibility is becoming more critical, conventional Web search engines still focus on searching for only popular and relevant information [3, 4]. In addition, they provide few clues to judge information credibility from Web search results. For instance, suppose that a user is searching for Web pages about *effective diet*. A lot of Web pages describe various diet methods. However, as in the case of the Atkins diet[1], Web pages containing suspicious information are often listed with

[1] Atkins diet: http://en.wikipedia.org/wiki/Atkins_diet



(a) Disputed sentence suggestion (b) Conventional keyword suggestion

Fig. 1. (a) Disputed sentence suggestion and (b) conventional keyword suggestion .

high rankings as popular and relevant Web search results, although there exists counter evidence against the information. Conventional Web search results provide only titles, a snippet, and URLs. This makes it difficult for users to check the credibility of such Web pages and obtain credible Web information using the conventional Web search engines. In worst cases, if users are not skeptical about Web pages, they can be mislead by incorrect information without knowing it. Therefore, automatic tools for alerting users and helping them judge Web information credibility are becoming increasingly necessary.

In this paper, to support users' credibility judgment in the Web search process, we propose a novel type of query suggestion to make users aware of suspicious Web information. When users issue queries to search for Web pages, our proposed system suggests some disputed sentences about the given queries. Fig.1-(a) and (b) show the case where a user issues keyword query *effective diet* to our Web search engine and to a conventional Web search engine, respectively. For this query, the conventional Web search engine suggested keywords *effective diet plans*, *effective atkins diet*, and so on, in Fig.1(b). As this example shows, conventional query suggestion techniques focus on supporting users in making their intent clearer and making it easy to search for Web pages matching their intent. In addition, suggested queries are often keywords. On the other hand, our implementation suggested that some phrases such as *"diet is an effective treatment"* and *"the atkins diet is not any more effective than other diet strategies"*, were disputed by some Web pages, in Fig.1(a). We think that disputed sentence suggestion has the following advantages over conventional keyword query suggestion techniques for supporting users' credibility judgment in the Web search process:

- Users can clearly recognize the existence of suspicious statements on the Web even if they are not overly concerned about the credibility of Web information.
- Users can find some queries to collect clues for credibility judgment of suspicious information from the viewpoint of counter evidence.
- Users can intuitively understand the meaning of disputed statements as opposed to just reading a list of keywords.

The remainder of the paper is organized as follows. In the next section, we discuss related work. Section 3 describes our methods for collecting disputed sentences from the Web and ranking them. In Sections 4, we show the effectiveness of our system for supporting users' credibility judgment in the Web search process, comparing it to

some query suggestion techniques. The last section concludes the paper and outlines our future research directions.

2 Related Work

2.1 Query suggestion

There has been a lot of work on query suggestion, including query expansion [5] and query substitution [6, 7]. Cui et al. studied a method to expand given queries by considering the gap between document space and query space [5]. Boldi et al. proposed a method to classify users' query reformulation operations into *generalization*, *specialization*, *error correction*, or *parallel move* towards query substitution [6]. Kotov et al. addressed a framework to automatically generate potentially interesting questions from Web pages and suggest them in the Web search process, so that Web search engines can naturally lead the users to Web search results they want [7]. These approaches are very sophisticated. However, they still focus on only mismatching problems between users' intent and given queries to improve users' Web search experience. Even if some queries are suggested, it is still difficult to call users' attention to suspicious Web information or to support users in collecting clues for credibility judgment from the Web.

2.2 Measuring Web information credibility

Some researchers have developed methods to measure the credibility of specific Web contents from specific credibility aspects. Especially, research aimed at measuring the credibility of contents on social networks like Yahoo Answers[2] and Twitter[3] is becoming more popular. Suryanto et al. focused on the expertise of answerers on Q&A sites to evaluate answers posted on them [8]. Castillo et al. proposed a method to automatically judge the credibility level of news propagated through Twitter by analyzing tweets and re-postings about news [9]. To measure the credibility of Web pages and Web search results from various viewpoints, some researchers have developed prototype systems [10, 11]. These systems visualize scores of Web search results on the basis of important credibility criteria, such as popularity, content typicality, and content freshness. If users want to briefly filter out non-credible Web information, measuring the credibility of Web information can be useful.

2.3 Supporting users' credibility judgment on the Web

Some studies have focused on promoting and helping users judge credibility by themselves. Pirolli et al. used WIKIDASHBOARD, the system for visualizing edit histories of Wikipedia articles, to study how the system affects users' credibility judgments on those articles [12]. Ennals et al. developed DISPUTE FINDER, a system that highlights suspicious sentences on Web pages users are browsing [13]. DISPUTEFINDER can highlight only suspicious sentences in its database and it is not robust in a Web search process

[2] Yahoo Answers: http://answers.yahoo.com/
[3] Twitter: http://twitter.com/

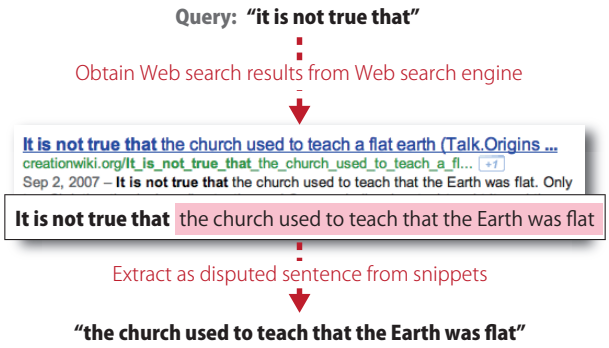


Fig. 2. Example of disputed sentence extraction from snippets of Web search results.

where different Web search results are shown for different queries. On the other hand, our system is robust. In addition, our system enables users to get an overview of suspicious statements for given queries. Some researchers have developed methods to find alternative statements for checking the credibility of suspicious statements [14, 15].

3 Disputed Sentence Suggestion

In this section, we describe our method for searching for disputed sentences about a given query on the Web and suggesting disputed sentences to users. Our system works as follows:

1. Collect disputed sentence candidates about a given query from the Web.
2. Score the disputed sentence candidates by considering the typicality of the disputed sentences and their relevancy to the given query.
3. Suggest top-k of the disputed sentences to users

3.1 Extracting disputed sentences from the Web

Detecting sentences that other documents dispute is quite hard because it requires deep natural language processing for large corpora. To bypass this problem, Ennals et al. proposed a method to collect disputed sentences from the Web by using linguistic patterns [16]. We apply their method to collect disputed sentences about a given query.

Ennals et al. prepared a set of dispute clue patterns that could indicate that a statement S is disputed, such as *"the misconception that S"* and *"it is not true that S"*. Then, as shown in Fig.2, they issued the prepared linguistic patterns as phrase queries into Web search engines. After that, they extracted the sentences that appeared just behind one of the patterns from obtained Web search results.

While Ennals et al. focus on collecting a large indefinite number of disputed sentences in batch processing, our goal is to collect disputed sentences related to a user's query in real time. To achieve this goal, we selected 15 of the 54 effective linguistic

Table 1. Linguistic patterns to collect disputed sentences from the Web

Linguistic pattern
"no proof that", *"mistakenly believe that"*, *"no evidence that"*, *"lie that"*
"it is not true that", *"it is not clear that"*, *"it is unlikely that"*, *"it is wrong that"*
"it is not to say that", *"it is denied that"*, *"into believing that"*
"misconception that", *"myth that"*, *"it is speculated that"*, *"it is delusion that"*

patterns that Ennals et al. suggested in their paper. The linguistic patterns are shown in Table 1. We implemented the method to collect disputed sentences by using these selected linguistic patterns. Our method works as follows:

1. Issue a given query and each of the prepared linguistic patterns into Yahoo Search Web API[4] (e.g. given keyword query *global warming cause*, the system issues expanded query *global warming cause AND "it is not true that"*).
2. Split snippets of Web search results into sentences using some delimiters (e.g. ".", ",", !, ?, *and*, *but*, and so on).
3. Extract sentences that appear just behind one of the prepared linguistic patterns from collected sentences as disputed sentences.
4. If the extracted sentences start with a specific stopword, such as a pronoun, a conjunction, a preposition, or a meaningless term, the system eliminates them from a list of the disputed sentences.

3.2 Ranking disputed sentences

After collecting disputed sentence candidates from Web search results about a given query, the system ranks the candidates to suggest useful disputed sentences to users.

Here we denote a given query and a disputed sentence by q and d, respectively. We estimate the importance of each collected disputed sentence $D = \{d_1, d_2, \ldots, d_n\}$, focusing on the relevance to given query q. Then we order them by using the following equations on the query likelihood model [3]:

$$p(d|q) \propto p(d) \prod_{t \in q} ((1-\lambda)p(t|M_c) + \lambda p(t|M_d)) \tag{1}$$

$$p(q|M_d) = \prod_{t \in q} \frac{tf_{t,d}}{L_d} \tag{2}$$

where M_d is the language model of disputed sentence d, $tf_{t,d}$ is the term frequency of term t in disputed sentence d, and L_d is the number of terms in disputed sentence d. λ is a weighting parameter and $0 < \lambda < 1$. M_c is the language model built from the entire disputed sentence collection D.

[4] Yahoo! Search Web API:
http://developer.yahoo.co.jp/webapi/search/

When using the query likelihood model, the prior probability of a document $p(d)$ is often treated as uniform across all d. However, we calculate $p(d)$ to consider how popular the collected disputed sentences are on the Web.

One way to determine the popularity of disputed sentences on the Web is to count how many times they appear. However, as exemplified by sentences *"a mobile phone is bad for your health"* and *"mobile phones are a health risk"*, the same statements are often expressed in different words. Therefore, it is difficult to measure the popularity or typicality of sentences by simply using their frequency. To solve this problem, we use the *LexRank* algorithm [17]. In this algorithm, a graph is created from text contents, where text contents are nodes and textual similarity between text nodes is the weight of the edge. Then the centrality of the text nodes is calculated by using the graph. The system measures $p(d)$ as the typicality of each disputed sentence $d \in D = \{d_1, d_2, \ldots, d_n\}$ using the LexRank algorithm. This computation is formalized as follows:

$$\mathbf{T} = \alpha \mathbf{S}^* \times \mathbf{T} + (1 - \alpha)\mathbf{p}, \quad \text{where } \mathbf{p} = \left[\frac{1}{n} \right]_{n \times 1} \tag{3}$$

Here, $\mathbf{S}^*$ is the column normalized adjacency matrix of the similarity matrix $\mathbf{S}$, where each $\mathrm{sim}(d_i, d_j)$ denotes the cosine similarity between disputed sentences d_i and d_j. $\mathbf{T}$ is the typicality score vector, where each $t(d)$ denotes the typicality of disputed sentence d on the Web. We approximate $p(d)$ by $t(d)$.

In practice, $p(d|q)$ of disputed sentence $d \in D$ is calculated as follows:

1. Collect disputed sentences $D = \{d_1, d_2, \ldots, d_n\}$ from the Web (as explained in Section 3.1).
2. Calculate $p(q|M_c)$ and $p(q|M_d)$ of all $d \in D$ using eq. 2. In this process, the system ignores stopwords in the collected disputed sentences.
3. Create feature vectors of the disputed sentences by ignoring stopwords. To weight terms of the vectors, the system uses tf-idf weighting.
4. Calculate typicality $t(d)$ as $p(d)$ of all disputed sentence $d \in D$ using 3. The system recursively calculates the typicality scores T of all disputed sentences D until the scores are converged.
5. Calculate $p(d|q)$ of all $d \in D$ using the results of step 2 and 4 on eq. (1).

4 Experiment

We conducted an experiment to examine how effective our method is for enhancing users' credibility judgment in the Web search process. In this experiment, we used the following three viewpoints to evaluate our disputed sentence suggestion and compared our method with two baseline methods:

- **Relevance:** How relevant are suggested disputed sentences to a given query?
- **Alertingness:** How critical will users become about Web search results for a given query when they look at suggested disputed sentences?
- **Usefulness:** How useful are suggested disputed sentences in searching for additional evidence to judge the credibility of Web search results for a given query?

We compared our method with two baseline methods. One is a conventional keyword query suggestion method, which suggests related queries based on users' query logs (called **CKS**). To imitate the conventional keyword query suggestion, we issued each of the prepared queries into Yahoo Web Search[5] and used its suggested keywords. As for ranking order of suggested queries, we simply used the order in which Yahoo Web Search suggested the queries. The other baseline method is the typicality-based disputed sentence suggestion method, which focuses on only $p(d)$ in the process we explained in Section 3 (called **TDS**). To execute disputed sentence suggestion, we set our method (called **Ours**) and **TDS** to collect 40 Web search results for each of the given queries. In addition, we set $\lambda = 0.5$ on eq. (1) for **Ours**.

4.1 Metrics for evaluation

To evalate the effectiveness of **Ours**, **TDS**, and **CKS** from the viewpoints of relevance, usefulness, and alertingness, we checked how many relevant/alerting/useful suggestions appeared in top-k rankings. To evaluate them, we used the following equations:

$$r@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Relevant(S_q(k))|}{k} \tag{4}$$

$$a@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Alerting(S_q(k))|}{k} \tag{5}$$

$$u@k = \frac{1}{|Q|} \sum_{q \in Q} \frac{|Useful(S_q(k))|}{k} \tag{6}$$

Here Q means a query set and $S_q(k)$ means top-k suggested keywords/sentences that each algorithm outputs for the input query q. $|S|$ indicates the size of the set S. $Relevant(S)$, $Alerting(S)$, and $Useful(S)$ respectively mean the suggested keywords/sentences in S judged as relevant, alerting, and useful by least two human evaluators.

4.2 Participants and materials

We recruited three human evaluators for this experiment. They all had experience in using Web search engines. For this experiment, we manually prepared 20 queries from five categories. The queries are shown in Table 2.

4.3 Procedure

In this experiment, we input each of the queries in Table 2 to the three methods (**Ours**, **TDS**, and **CKS**) and showed the suggested keywords/sentences to the evaluators. Then we asked them to evaluate the suggested keywords/sentences in terms of relevance, alertingness, and usefulness. Before the evaluators evaluated suggested keywords/sentences for each query, we showed them a brief description like:

[5] Yahoo Web Search: http://search.yahoo.com/ and then manually obtained suggested queries

Table 2. Query set.

Category	Query
Controversy	dinosaurs extinction, global warming, earthquake cause plastic recycling, renewable energy
Health	effective diet, cancer treatment, mobile phone health vaccine side effects, caffeine overdose
Misunderstanding	light bulb inventor, telephone inventor civil war reason, caesar salad name
Food	coffee health, grapefruit seed extract benefit, potato poison
Finance	forex risk, mortgage refinancing, personal debt reduction

Suppose that you input the keywords "mobile phone health" to a Web search engine, and then the search engine shows some Web search results. In addition to the search results, the search engine also suggests some keywords/sentences as highlighted below. Now you are concerned about the credibility of the Web search results the search engine returns for the input keywords "mobile phone health". On the basis of the suggested keywords/sentences, please answer the following three questions.

In each evaluation task for **CKS**, we showed the evaluators raw keyword queries suggested by Yahoo. In each evaluation task for **Ours** and **TDS**, we added the expression *"Some people doubt:"* to each of disputed sentences and showed the combined sentences to the evaluators (e.g., if the system found the disputed sentence *"mobile phone has a health risk"* for the query *"mobile phone health"*, the system suggested *"Some people doubt:x mobile phone has a health risk"* to the evaluators).

To evaluate the relevance, alertingness, and usefulness of each of the suggested keywords/sentences, we asked the evaluators to answer the following three questions with YES or No:

- **Relevance:** *Are suggested keywords/sentences related to a given query?*
- **Alertingness:** *Will you become more critical of Web search results about a given query when looking at suggested keywords/sentences?*
- **Usefulness:** *Are suggested keywords/sentences useful in searching for additional evidence to judge the credibility of Web search results about a given query?*

4.4 Results

Fig.3 shows the averages of relevance scores, usefulness scores, and alertingness scores for each method on a query set, using equations $r@k$, $a@k$, and $u@k$. According to Fig.3(a), in terms of relevance, **CKS** performed much better than **Ours** and **TDS**. However, **Ours** was not so bad at suggesting relevant sentences ($r@1 = 0.80$, $r@3 = 0.70$, and $r@5 = 0.69$). **CKS** is based on query logs a lot of users issued in actual Web search sessions. Therefore, **CKS** can precisely suggest relevant keywords for given queries, unlike **Ours** and **TDS**, which analyze documents to suggest disputed sentences. The relevance performance of **Ours** was higher than **TDS** on $r@1$, $r@3$, and $r@5$. Especially, according to $r@1$, **Ours** was higher by 33% than **TDS**. We think that query

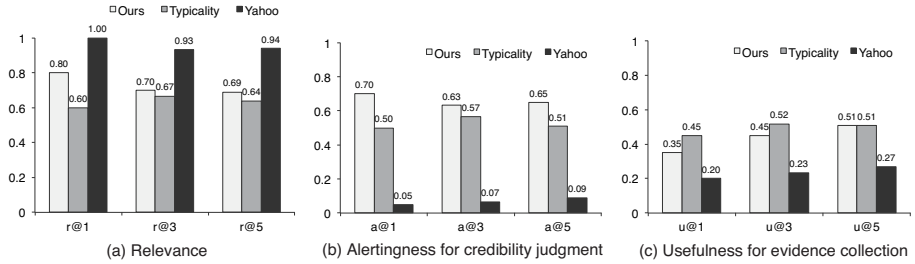


Fig. 3. Relevance, alertingness, and usefulness of three methods.

likelihood $\prod_{t \in q}((1-\lambda)p(t|M_c) + \lambda p(t|M_d))$ in eq. (1) contributes to the relevance score.

In terms of alertingness, Fig.3(b) shows that with **Ours** and **TDS**, at least 50% of suggested sentences in the top 5 rankings called the evaluators' attention to credibility judgment. The performance of **Ours** was higher than that of **TDS** on $a@1$, $a@3$, and $a@5$. On the other hand, as the result for **CKS** indicates, the evaluators did not become more careful of their credibility judgment at all, even when looking at the keywords suggested ($a@1$=0.05, $a@3$=0.07, $a@5$=0.09). This result is natural because the objective of conventional keyword suggestion is different from that of our disputed sentence suggestion, which aims at alerting users to suspicious information.

Fig.3(c) shows that the two methods that suggest disputed sentences performed better than **CKS**. This indicates that the disputed sentence suggestion can be more useful in searching for additional clues for credibility judgment on Web pages about given queries than conventional keyword suggestion.

Before this experiment, we expected that disputed sentence suggestion would be more useful in searching for clues for credibility judgment. However, neither of the two disputed sentence suggestion methods showed show usefulness scores as high as we expected. In addition, $u@1$ and $u@3$ of **Ours** were less than those of **TDS** (scores of the two methods were the same on $u@5$). We interviewed the evaluators after the experiment. Below is a typical evaluator comment:

> *The suggested disputed phrases themselves were often useful as evidence or clues to judge the credibility of Web pages about given queries. So I judged that I didn't need to search for Web pages using the suggested disputed phrases.*

This indicates that disputed sentence suggestion was not necessarily useless, although Fig.3(c) indicate that it was not so useful in searching for additional credibility evidence from the Web. We think that if systems can suggest disputed sentences relevant to given queries, the disputed sentences themselves are likely to be useful clues for credibility judgment on Web pages for given queries. The experimental results shows that **Ours** suggested more relevant disputed sentences than **TDS**

Finally, these experimental results indicate that, by suggesting disputed sentences with reasonable relevance to given queries, our proposed method can make users more critical in the Web search process than the baseline methods can.

Table 3. Good examples of suggested disputed sentences. Numbers in parentheses are ranking orders of suggested disputed sentences.

Query	Suggested disputed sentence
potato poison	*"green potatoes are poisonous"* (1)
vaccine side effects	*"tamiflu does not have potential side effects"* (5)
caesar salad name	*"the salad takes its name from julius caesar"* (1)
civil war reason	*"slavery was the cause of the civil war"* (1)
earthquake cause	*"fracking causes earthquakes"* (3)

Table 4. Bad examples of suggested disputed sentences and error types. Numbers in parentheses are ranking orders of suggested disputed sentences.

Query	Suggested disputed sentence	Error type
potato poison	*"potatoes are fattening"* (5)	Non-relevant
global warming	*"global warming was the cause"* (1)	ambiguous
telephone inventor	*"agb was not the inventor"* (3)	Difficult to understand
plastic recycling	*"shipping plastics for recycling"* (2)	Grammatical error

4.5 Case study on disputed sentence suggestion

Table 3 and 4 show some examples of good and bad disputed sentences the system suggested for given queries using our proposed method. We manually checked the suggested disputed sentences the evaluators judged as bad and categorized them into four classes: *non-relevant*, *ambiguous*, *difficult to understand*, and *grammatical error*.

As evidenced by sentences like *"potatoes are fattening"* for query *potato poison*, the system sometimes suggested non-relevant disputed sentences to given queries. We think this is because even if extracted disputed sentences are not very relevant to given queries, our system gives them high scores if they are typical on the Web. We have to think about a way to balance query likelihood scores and typicality scores of disputed sentences on eq. (1).

Even when the system maintained the relevance of disputed sentence suggestion, ambiguous disputed sentences were sometimes suggested, such as *"global warming was the cause"* for query *global warming*. In our method, the system simply extracts disputed sentences that appear behind linguistic patterns like *"it is not true that"* and *"no proof that"*. Therefore, the system often suggests that ambiguous and difficult-to-understand disputed sentences without context be extracted from Web search results.

The evaluators misunderstood that the system suggested some meaningless disputed sentences for given queries, although the disputed sentences were actually relevant to the queries and alerting. One possible reason is that the evaluators could not understand the meaning of some terms in the disputed sentences because they were not familiar with the sentences. For instance, the system suggested disputed sentence *"agb was not the inventor"* for query *telephone inventor*. If users do not know that term *agb* means *Alexander Graham Bell*, they may think the disputed sentence is useless and then will not become more critical of Web pages for the given query. It is important to consider the comprehensibility of sentences when suggesting them to users.

The system also sometimes suggested disputed sentences that were grammatically difficult to read like *"someone doubts: shipping plastics for recycling"* for query *plastic recycling*. Our system splits sentences with punctuation marks and extracts disputed sentences by simply focusing on declarative content clauses (*that clauses*). Therefore, the system often fails to extract grammatically complete sentences. We need deeper natural language processing to handle this problem.

5 Conclusion

In this paper, we addressed a new type of query suggestion to call users' attention to credibility judgment of Web search results for given queries in the Web search process. Given a query, such as *mobile phone health*, our system searches for statements some Web pages dispute about the query and suggests ones like *"mobile phone are bad for your health"* to users. While conventional keyword query suggestion focuses on making it easy for users to search for Web pages that match their search intent, our disputed sentence suggestion focuses on making them aware of the existence of suspicious information or to get clues for a credibility judgment on Web pages about their input query. Even if users are not concerned with Web information credibility and do not know about the existence of suspicious Web information, they can notice suspicious information and become more critical in searching for credible Web pages.

The experiment results show our method suggested effective disputed sentences to help users more critically search for credible Web pages, although the method had lower performance in terms of relevancy than conventional keyword suggestion. To improve our disputed sentence suggestion, we first have to improve the relevance of suggested disputed sentences by optimizing the balance between the query likelihood scores and typicality scores of them. In addition, we need to think about a way to suggest comprehensible disputed sentences because sometimes users cannot understand the meaning of suggested disputed sentences even if they are relevant and typical to given queries. Another important issue is how to support users' search for credible Web pages efficiently using suggested disputed sentences.

To safely and efficiently obtain information from the vast Web, search systems focusing on credibility will become more important in the future . We believe our proposed system can contribute to credibility-oriented Web search.

Acknowledgment

The author wishes to thank Xiaojing Zhang, Yosuke Ishihara, and Daisuke Kitayama for their valuable comments. This work was supported by Grants-in-Aid for Scientific Research (No. 18049041) from MEXT of Japan, a Kyoto University GCOE Program entitled "Informatics Education and Research for Knowledge-Circulating Society".

References

1. Sillence, E., Briggs, P., Fishwick, L., Harris, P.: Trust and Mistrust of Online Health Sites. In: Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 2004). (2004) 663–670

2. Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S., Tanaka, K.: Trustworthiness Analysis of Web Search Results. In: Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007). (2007) 38–49
3. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 1998). (1998) 275–281
4. Liu, Y., Gao, B., Liu, T.Y., Zhang, Y., Ma, Z., He, S., Li, H.: BrowseRank: Letting Web Users Vote for Page Importance. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2008). (2008) 451–458
5. Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y.: Probabilistic Query Expansion Using Query Logs. In: Proceedings of the 11th international conference on World Wide Web (WWW 2002). (2002) 325–332
6. Boldi, P., Bonchi, F., Castillo, C., Vigna, S.: From "Dango" to "Japanese Cakes": Query Reformulation Models and Patterns. In: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI 2009). (2009) 183–190
7. Kotov, A., Zhai, C.: Towards natural question guided search. In: Proceedings of the 19th international conference on World wide web (WWW 2010). (2010) 541–550
8. Suryanto, M.A., Lim, E.P., Sun, A., Chiang, R.H.L.: Quality-Aware Collaborative Question Answering: Methods and Evaluation. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM 2009). (2009) 142–151
9. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web (WWW 2011). (2011) 675–684
10. Schwarz, J., Morris, M.: Augmenting web pages and search results to support credibility assessment. In: Proceedings of the 2011 annual conference on Human factors in computing systems (CHI 2011). (2011) 1245–1254
11. Yamamoto, Y., Tanaka, K.: Enhancing Credibility Judgment of Web Search Results. In: Proceedings of the 2011 annual conference on Human factors in computing systems (CHI 2011). (2011) 1235–1244
12. Pirolli, P., Wollny, E., Suh, B.: So You Know You're Getting the Best Possible Information: A Tool that Increases Wikipedia Credibility. In: Proceedings of the 27th international conference on Human factors in computing systems (CHI 2009). (2009) 1505–1508
13. Ennals, R., Trushkowsky, B., Agosta, J.M.: Highlighting Disputed Claims on the Web. In: Proceedings of the 19th international conference on World wide web (WWW 2010). (2010) 341–350
14. Kawahara, D., Inui, K., Kurohashi, S.: Identifying contradictory and contrastive relations between statements to outline web information on a given topic. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). (2010) 534–542
15. Yamamoto, Y., Tezuka, T., Jatowt, A., Tanaka, K.: Supporting Judgment of Fact Trustworthiness Considering Temporal and Sentimental Aspects. In: Proceedings of the 9th international conference on Web Information Systems Engineering (WISE2008). (2008) 206–220
16. Ennals, R., Byler, D., Agosta, J.M., Rosario, B.: What is disputed on the web? In: Proceedings of the 4th workshop on Information credibility (WICOW 2010). (2010) 67–74
17. Erkan, G., Radev, D.: LexRank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research **22** (2004) 457–479