A Method for Automatically Generating a Mediatory Summary to Verify Credibility of Information on the Web

Hideyuki Shibuki and **Takahiro Nagai** and **Masahiro Nakano**
Rintaro Miyazaki and **Madoka Ishioroshi** and **Tatsunori Mori**
Graduate School of Environment and Information Sciences,
Yokohama National University
{shib, nagadon, nakano, rintaro, ishioroshi, mori}@forest.eis.ynu.ac.jp

Abstract

In this paper, we propose a method for mediatory summarization, which is a novel technique for facilitating users' assessments of the credibility of information on the Web. A mediatory summary is generated by extracting a passage from Web documents; this summary is generated on the basis of its relevance to a given query, fairness, and density of keywords, which are features of the summaries constructed to determine the credibility of information on the Web. We demonstrate the effectiveness of the generated mediatory summary in comparison with the summaries of Web documents produced by Web search engines.

1 Introduction

Many pages on the Web contain incorrect or unverifiable information. Therefore, there is a growing demand for technologies that can enable us to obtain reliable information. However, it would be almost impossible to automatically ascertain the accuracy of information presented on the Web. Hence, the second-best approach is the development of a supporting method for judging the credibility of information on the Web.

Presently, when we wish to judge the credibility of information on the Web, we often read some relevant Web documents retrieved via Web search engines. However, Web search engines do not provide any suggestions in cases where the content of some documents conflicts with the content of other documents. Furthermore, the retrieved documents are too many to read and may not be ranked according to the credibility of the information they provide. In other words, information retrieval is not sufficient to support users' assessments of the credibility of information, and therefore, additional techniques are required for the same.

Several previous researches have been conducted for developing such techniques. Juffinger et al. (2009) ranked blogs in terms of their concurrence with well-verified information from sources such as a news corpus. Miyazaki et al. (2009) devised a method for extracting the description of the information sender, namely, the person or organization providing texts in Web pages. Ohshima et al. (2009) proposed a method for reranking Web pages according to users' regionality, which depends on two factors: uniformity and proximity. While the abovementioned studies mainly facilitate users' assessments of the credibility of individual Web pages, the following studies deal with the issues of credibility of information in multiple documents on the Web. Murakami et al. (2009) proposed a method to analyze semantic relationships such as agreement, conflict, or evidence between texts on the Web. Kawahara et al. (2009) reported a method for presenting overviews of evaluative information such as positive/negative opinions.

Although the above techniques facilitate users' assessments of credibility of information on the Web, there is still room for methods that support users' judgment. For example, when the truth of a statement[1] "Diesel engines are harmful to the environment" is to

---

[1] In this paper, a *statement* is defined as text such as an opinion, evaluation, or objective fact.

Query: Are diesel engines harmful to the environment?

Positive Opinion   **Negative Opinion**

YES, diesel is polluting the environment very much and release carbon monoxide.   **VS.**   Nope, they emit black smoke, but are not hazardous to environment, unlike petroleum.

The above sentences appear to contradict each other at first glance.
Read the following text, and judge whether they can coexist under the situation.

Guidance for Interpretation

Americans continue to perceive diesel as a "dirty" fuel, though today that image is only partly deserved. Because of their lower per-mile fuel consumption, diesel engines generally release less **carbon dioxide** - the heat-tapping gas primarily responsible for global warming - from the tailpipe. So that's a check on the good side of the pollution chart. But when it comes to **smog-forming pollutants and toxic particulate matter**, also known as soot, today's diesels are still a lot dirtier than the average gasoline car.

Comment: They may describe different types of "the environment."
Do the terms "smog-forming pollutants and toxic particulate matter" and "carbon dioxide" describe the same thing?
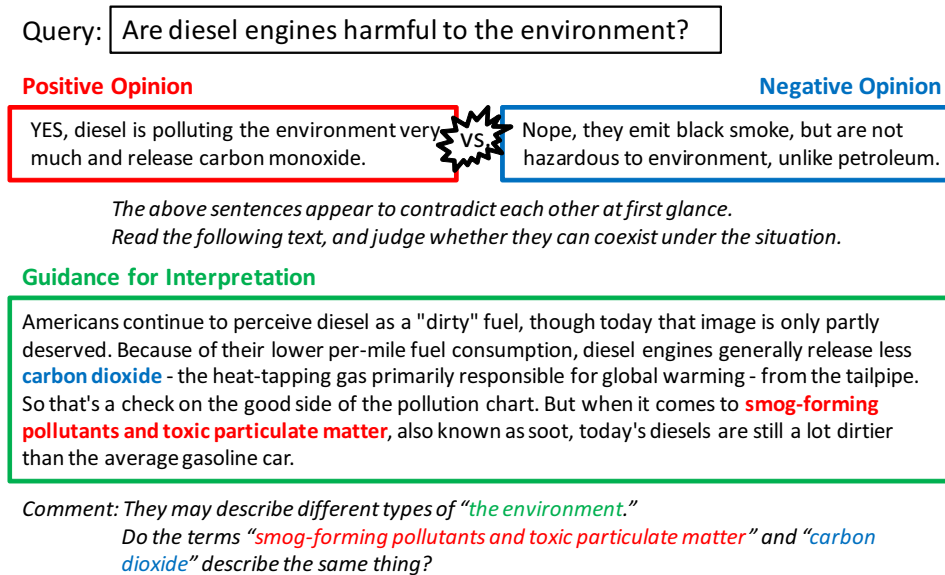
Figure 1: An example of the mediatory summary.

be verified, the following two contradictory groups of Web documents are obtained: one stating that "Diesel engines are harmful to the environment" and the other stating that "Diesel engines are not harmful to the environment." How does one resolve this conflict? Are the contents of one group that contains a less reasonable description wrong? On the other hand, if contents of both these groups are correct, then why do they appear to be contradictory to each other? A display of only the overview of statements in Web documents and the relationships between these statements does not always provide sufficient information to answer these questions.

In order to direct users to a reasonable interpretation written by one author, Kaneko et al. (2009) proposed the notion of a mediatory summary for *pseudo conflicts*, which are relationships between statements that appear to contradict each other at first glance but can coexist under a certain situation. However, Kaneko et al. did not describe any algorithms for automatically generating the mediatory summary. Therefore, in this paper, we propose a method for automatically generating the mediatory summary and demonstrate the effectiveness of generated mediatory summaries.

The rest of this paper is organized as follows. In Section 2, we describe the concept of a mediatory summary and the features required for automatically generating it. In Section 3, we describe an algorithm for generation of the mediatory summary. In Section 4, we present experimental results to demonstrate the effectiveness of the generated mediatory summary. Section 5 provides the conclusion.

2 Mediatory Summary

The generation of the mediatory summary is a type of informative summarization based on passage extraction. Figure 1 shows an example of the ideal mediatory summary for the query "Are diesel engines harmful to the environment?" The text in boxes with thick lines is extracted from Web documents, and the italicized text is generated through templates. The user is shown both positive and negative responses to the query and appropriately guided on how to interpret them. One of the most difficult issues in the automated generation of the mediatory summary is the extraction of the most suitable passage that is used for interpretation.[2]

---

[2]Owing to space limitations, we have omitted the discussion on generation through templetes.

Nakano et al. (2010) constructed a text summarization corpus for determining the credibility of information on the Web. Their corpus contains six query statements, the Web document collections retrieved for each query, and 24 summaries made by four persons per query. We analyzed these summaries from the viewpoint of mediatory summary generation and observed that mediatory summaries usually display the following three features.

The first feature is the high relevance to a given query. We can approximately determine whether the text displays this feature by examining whether or not it contains content words in the query such as "diesel," as in Figure 1. The second feature is fairness, or in other words, evenly describing both positive and negative opinions. We can approximately determine whether the text displays this feature by examining whether or not it contains words for both opinions and having different, typically opposite meanings such as "lot" and "less," as in Figure 1. It should be noted that words with opposite meanings are not limited to antonyms. In the case of the query "Are diesel engines harmful to the environment?," "carbon dioxide" and "smog-forming pollutants" should be also regarded as words with opposite meanings. The third feature is the high density of words of the above two types in a text. In addition to appropriateness of a text with high density as a summary, such a text is likely to be a text contrasting both sides of contents.

3 Proposed Method

3.1 Outline

We propose a method for generating mediatory summaries by extracting passages that display the features described in the previous section. First, we define the content words related to the topic of a query as *topic keywords*.[3] Next, we define the content words cor-

---

[3]Although topic words are not restricted to words in the given query, we use only the words in the given query as topic words in this paper. Inclusion of words other than those in the given query is a topic for future research.
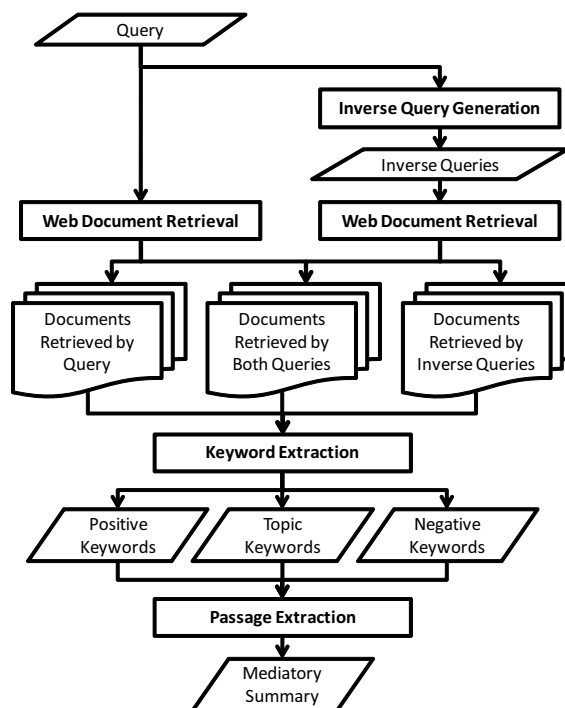

Figure 2: Outline of the proposed method.

responding to the positive and negative opinions as *positive keywords* and *negative keywords*, respectively. Although topic keywords appear in the given query, positive/negative keywords hardly appear in the given query. Therefore, positive/negative keywords have to be extracted from other text.

Figure 2 shows the outline of the proposed method. First, in order to find contents opposed to a given query, *inverse queries* are generated. We define an inverse query as a query generated by replacing a word in the given query with its antonym. Next, the given query and the inverse queries are used to retain three sets of Web documents, which are likely to contain more positive keywords, negative keywords, and neither. The topic, positive, and negative keywords are then extracted from these sets. Finally, the extracted keywords are used to extract passages from the three sets of Web documents; these passages are ranked in order of the score described in Section 3.5 as a mediatory summary.

3.2 Inverse Query Generation

Inverse queries are generated by simply replacing a word in the given query with an antonym of the word using a dictionary for antonyms. For example, if the query is "Is safety of LASIK operation high?," and if "risk" and "low" are input into the dictionary as antonyms of "safety" and "high," respectively, then the two generated inverse queries are "Is risk of LASIK operation high?" and "Is safety of LASIK operation low?" Positive keywords are words in the given query that are to be replaced with their opposite words in the inverse queries. On the other hand, the newly introduced opposite words in the inverse queries are regarded as negative keywords. In the above example, "safety" and "high" are positive keywords, whereas "risk" and "low" are negative ones. If there is no replaceable word, the inverse query is not generated.

3.3 Web Document Retrieval

Using a given query and the corresponding inverse queries, Web documents are retrieved via TSUBAKI (Shinzato et al., 2008); TSUBAKI is an open-search engine with an infrastructure based on deep Japanese natural language processing, and it can accept a natural-language sentence as a query. The number of documents retrieved per query is tentatively set to 100 in our experiment described in Section 4.

The retrieved documents are classified into the following three document sets: one containing documents retrieved by the given query but not retrieved by the inverse queries, one containing documents retrieved by the inverse queries but not retrieved by the given query, and one containing documents retrieved by both the given query and the inverse queries. These three document sets are termed D_{query}, $D_{inverse}$, and D_{both}, respectively. Words appearing frequently in D_{query} are likely to be positive keywords, and those appearing frequently in $D_{inverse}$, negative keywords.

TSUBAKI has an optional function for au-

Table 1: An example of extracted keywords.

word	rank tf	rank POS	rank NEG	polarity
LASIK	1	1	1	other
operation	2	2	2	other
eyesight	3	3	3	other
examination	19	13	60	positive
glasses	47	58	75	other
blindness	61	5,045	20	negative
effect	66	72	111	positive
complications	77	206	58	negative

tomatically expanding a submitted query by including the negative form of the main verb in the query. For example, if "Is safety of LASIK operation high?" is the submitted query, then the expanded query is "Is safety of LASIK operation not high?" The documents retrieved for the expanded query derived from the given query are regarded as documents retrieved for inverse queries. Similarly, the documents retrieved for the expanded query derived from the inverse queries are regarded as documents retrieved for the given query. Hence, even if there is no inverse query, possibly conflicting documents can be retrieved.

3.4 Keyword Extraction

Positive and negative keywords are extracted from the retrieved Web documents as follows. First, the positive score $sc_{POS}(w)$ and negative score $sc_{NEG}(w)$ of a word w are calculated by Equations (1) and (2), respectively:

$$sc_{POS}(w) = \frac{df(w, D_{query}) \cdot tf(w)}{df(w, D_{inverse}) + 1} \quad (1)$$

$$sc_{NEG}(w) = \frac{df(w, D_{inverse}) \cdot tf(w)}{df(w, D_{query}) + 1} \quad (2)$$

where $tf(w)$ is the frequency of w in all retrieved documents, and $df(w, D)$ is the number of documents containing w in D. The $sc_{POS}(w)$ is higher and $sc_{NEG}(w)$ is lower if the word w appears more frequently in the documents retrieved by the given query and less frequently in the documents retrieved by the inverse queries. The frequency $tf(w)$ is

used to express these scores in order to consider the global importance of w in the entire retrieved document set.

Because words such as "LASIK" in the query "Is safety of LASIK operation high?" have high scores in terms of both $sc_{POS}(w)$ and $sc_{NEG}(w)$, such words should not be extracted as positive or negative keywords. Because the number of documents in D_{query} is different from that in $D_{inverse}$, $sc_{POS}(w)$ cannot be directly compared with $sc_{NEG}(w)$. Hence, $sc_{POS}(w)$ and $sc_{NEG}(w)$ are normalized, and $rank_{POS}(w)$ and $rank_{NEG}(w)$ are compared. The ranking functions $rank_{POS}(w)$ and $rank_{NEG}(w)$ are defined as the nth place ranks when all words are ranked in the descending order of $sc_{POS}(w)$ and $sc_{NEG}(w)$, respectively. We consider the top C_{rank} words on the $tf()$ ranking as candidates for possible keywords. If $rank_{POS}(w) - rank_{NEG}(w)$ or $rank_{NEG}(w) - rank_{POS}(w)$ is greater than C_{dif}, then we regard w as a positive or negative keyword, respectively. Table 1 shows an example of the extracted keywords when the given query is "Is safety of LASIK operation high?" In this paper, C_{rank} and C_{dif} are tentatively set to 100 and 20, respectively, in a preliminary experiment using several queries except the ones described in Section 4.[4] Finally, the positive/negative keywords, mentioned in Section 3.2, are respectively added to the positive/negative keyword sets described above.

The topic keywords are the words in the given query excluding the positive/negative keywords.

3.5 Passage Extraction

Passages suitable for a mediatory summary are extracted through the following four stages.

For all sentences in the document sets described in Section 3.3, the first stage involves the recognition of sentences that are useless for mediatory summary. We regard insuffi-

---

cient or incomplete sentences as useless sentences. We simply consider a sentence to be sufficient when the sentence has, at least, one verb phrase and one noun phrase and when the sentence contains more than two verb/noun phrases. We simply consider a sentence to be insufficient if it is not a sufficient sentence. When the expression "...," which indicates an omission, appears at the end of a sentence, then we consider the sentence incomplete. This recognition of sentences plays an important role in the calculation of scores in subsequent stages.

The second stage involves the calculation of the score of each sentence. When KW is defined as a set of all the topic, positive, and negative keywords, the basic score $sc_{BAS}(s)$ of sentence s is calculated by Equation (3).

$$sc_{BAS}(s) = \frac{\sum_{w \in KW} appear(w, s)}{|KW|} \qquad (3)$$

where $appear(w, s)$ is a function whose value is 1 if w appears in s and 0 otherwise. If s contains many different keywords, $sc_{BAS}(s)$ acquires a high value. If s is recognized as a useless sentence, then $sc_{BAS}(s)$ is multiplied by $C_{useless}$ as a penalty. In this paper, $C_{useless}$ is tentatively set to 0.5 in the case of ungrammatical sentences and 0 in the case that the end of the sentences is omitted. When the score of a sentence is calculated, the fairness described in Section 2 is determined in the following manner. If s contains positive/negative keywords besides topic keywords, the $sc_{BAS}(s)$ is multiplied by C_{basic}. The negative form of positive/negative keywords is considered, and C_{basic} is tentatively set to two if either a positive or negative expression appears in s and to three if both positive and negative expressions appear in s.

The third stage involves the application of a smoothing method to raw scores of a sentence in order to suppress over-fragmentation of passages. As given in Equation (4), the smoothed score $sc_{SMO}(s_i)$ for the ith sentence s_i in a document is calculated using the Hann

function, whose window length is L.

$$sc_{SMO}(s_i) = \sum_{j=-\frac{L}{2}}^{\frac{L}{2}} (sc_{BAS}(s_{i+j}) \cdot hf(j)) \quad (4)$$

$$hf(k) = 0.5 + 0.5 \cos 2\pi \frac{k}{L} \quad (5)$$

The value of L is tentatively set as 5 in this study. Insufficient sentences may convey useful information to readers when they are embedded in an appropriate context. On the other hand, incomplete sentences do not. Therefore, $sc_{SMO}(s)$ of such omitted sentences is set as 0 even after smoothing. If all types of keywords appear in the Hann window, then $sc_{SMO}(s)$ is multiplied by C_{smooth} because a passage containing s is likely to be a part of a mediatory summary. The value of C_{smooth} is tentatively set as 2 in this study.

The fourth stage involves the extraction and ranking of passages. Every series of sentences with $sc_{SMO}(s)$ greater than $\frac{1}{N}$ of the maximum score in the document is extracted as a passage. N is tentatively set as 3. The score $sc_{PAS}(p)$ of passage p is the highest score $sc_{SMO}(s)$ of sentence s in the passage. If all types of keywords appear in p, the $sc_{PAS}(p)$ is multiplied by $C_{passage}$, as in the third stage. Note that the length of the extracted passages is not set to L sentences and that passages that contain sentences multiplied by C_{smooth} are not always multiplied by $C_{passage}$. $C_{passage}$ is tentatively set as 3 in this study. Because of summarization, the passages whose lengths are nearer to the ideal length C_{length} are ranked higher. The final score $sc_{FIN}(p)$ is calculated by Equation (6).

$$sc_{FIN}(p) = \exp(sc_{PAS}(p) - \alpha \cdot er(p)) \quad (6)$$

$$er(p) = |C_{length} - nc(p)| \quad (7)$$

where $nc(p)$ is the number of characters in p, and α is a coefficient. C_{length} and α are tentatively set to 300 and 0.02, respectively.

4 Experiment

4.1 Conditions

Because the proposed method is the first method for automated generation of media-

tory summary, there is no existing method to directly compare our proposed method with. Therefore, we compare the proposed method with the following three methods. The first method, $KWtf$, uses frequent words instead of the three types of keywords described in Section 3.4. The second method, Lin_{TSU}, uses an existing summarization module for summarizing the top documents in order of the score in which TSUBAKI retrieves them. The third method, Lin_{scFIN}, uses the same existing summarization module for summarizing documents containing the top passages in order of the score $sc_{FIN}()$ described in Section 3.5. We employ Lingua::JA::Summarize::Extract[5] as a summarization module, which extracts sentences containing more characteristic words; this extraction is based on the word frequency and word bigram frequency in a given document.

$KWtf$ is compared with the proposed method in order to investigate the effectiveness of keywords in terms of polarity. One of the functions of the proposed method is the classification of the top C_{rank} words on the $tf()$ ranking into positive keywords, negative keywords, and others in order to determine the fairness described in Section 2. $KWtf$ is simply based on frequent words and does not classify them. In other words, all of the top C_{rank} words on the $tf()$ ranking are used as keywords without polarity. It should be noted that no rewards can be obtained using C_{basic}, C_{smooth}, and $C_{passage}$ in the passage extraction described in Section 3.5, although penalties can be obtained using $C_{useless}$ and C_{length}.

Lin_{TSU} is compared with the proposed method in order to clarify the difference between the summarization by our method and that by a method used for general purposes. In other words, we investigate whether or not the extraction of sentences containing more characteristic words is sufficient for generating mediatory summaries.

Lin_{scFIN} is compared with the proposed method in order to investigate the appropri-

---

[5]http://search.cpan.org/~yappo/Lingua-JA-Summarize-Extract-0.02/

Table 2: Average precision of appropriateness of summaries generated by each query.

	Top 3	Top 5	Top 10		Top 3	Top 5	Top 10
Are diesel engines harmful to the environment?				Is safety of LASIK operation high?			
Proposed	100.0%	60.0%	36.7%	Proposed	33.3%	20.0%	20.0%
KWtf	0.0%	0.0%	0.0%	KWtf	0.0%	0.0%	0.0%
Lin$_{TSU}$	66.7%	40.0%	26.7%	Lin$_{TSU}$	33.3%	20.0%	30.0%
Lin$_{scFIN}$	0.0%	13.3%	16.7%	Lin$_{scFIN}$	0.0%	0.0%	0.0%
Are whales endangered species?				Does asbestos have toxics?			
Proposed	55.6%	33.3%	30.0%	Proposed	33.3%	40.0%	56.7%
KWtf	0.0%	0.0%	0.0%	KWtf	33.3%	20.0%	30.0%
Lin$_{TSU}$	11.1%	20.0%	13.3%	Lin$_{TSU}$	33.3%	20.0%	30.0%
Lin$_{scFIN}$	22.2%	13.3%	10.0%	Lin$_{scFIN}$	0.0%	20.0%	30.0%
Is catch and release a better way of fishing?				Does carbon dioxide cause global warming?			
Proposed	33.3%	26.7%	13.3%	Proposed	0.0%	0.0%	6.7%
KWtf	0.0%	0.0%	6.7%	KWtf	0.0%	0.0%	0.0%
Lin$_{TSU}$	66.7%	46.7%	26.7%	Lin$_{TSU}$	11.1%	6.7%	3.3%
Lin$_{scFIN}$	33.3%	26.7%	13.3%	Lin$_{scFIN}$	0.0%	0.0%	13.3%

ateness of the extracted passages. Another function of the proposed method is the reranking of passages regardless of the order of documents containing the passages during document retrieval. Therefore, a set of documents summarized by the proposed method may be different from the set of documents summarized by Lin_{TSU}. Therefore, it is observed that Lin_{scFIN} handles the same documents as the proposed method.

Because the mediatory summary is a novel concept, methods for evaluating it have not been developed yet. Although ROUGE (Lin and Hovy, 2003) is one of the most popular methods for evaluation of summaries, it may not be appropriate for the evaluation of the mediatory summary because the scoring based on N-gram in this method cannot be used to consider the fairness described in Section 2. Therefore, we evaluate the methods through the binary judgment of three human assessors, If the top summaries produced by each method are deemed to be appropriate by the three human assessors, we will be able to facilitate users' assessments of the contradictory opinions that are relevant to the given query.

Because we consider that filling a passage with all the information necessary to facili-

tate users' assessments is more important than shortening the passage under conditions for generation of mediatory summary, we imposed no limitation on the length of passages but the resultant penalty is obtained using Equations (6) and (7). The average length of all summaries generated by the proposed method and $KWtf$ was 288.4 characters, and none of the summaries exceeded 500 characters. Therefore, we allowed Lin_{TSU} and Lin_{scFIN} to generate summaries as long as 500 characters, which is about 200 characters longer than summaries generated by the proposed method and $KWtf$. We instructed the assessors to not judge the appropriateness of the summaries on the basis of their length.

For the experiment, we prepared the following six queries: "Are diesel engines harmful to the environment?," "Is safety of LASIK operation high?," "Are whales endangered species?," "Does asbestos have toxics?," "Is catch and release a better way of fishing?," and "Does carbon dioxide cause global warming?" We used the Japanese morphological analyzer, MeCab.[6]

---

[6]http://mecab.sourceforge.net/ (in Japanese)

Table 3: Average precision of assessors in terms of appropriateness of overall summaries.

	Top 3	Top 5	Top 10
Proposed	42.6%	30.0%	27.2%
KWtf	5.6%	3.3%	6.1%
Lin$_{TSU}$	37.0%	25.6%	21.7%
Lin$_{scFIN}$	9.3%	12.2%	13.9%

4.2 Result and Consideration

The kappa values between each pair of assessors' judgments on the appropriateness of the summaries were 0.79, 0.77, and 0.76, respectively; these values indicate a high level of agreement among assessors' judgments.

Table 2 shows the average precision[7] of the assessors in terms of appropriateness of summaries on the basis of responses to each query, and Table 3 shows the overall precision. The columns in Tables 2 and 3 show the precision of appropriateness for the top 3, top 5, and top 10 summaries produced by each method. It should be noted that there are only a few passages suitable for mediatory summary in all of the retrieved documents, and therefore a method for placing such suitable passages at a higher rank is more effective. We confirmed that the proposed method provided the best overall results among all the compared methods.

The difference between the proposed method and KWtf shows that classification of frequent words into positive keywords, negative ones, and others, in other words, the fairness described in Section 2 contributed to generation of appropriate mediatory summaries.

The difference between Lin$_{TSU}$ and Lin$_{scFIN}$ indicates that the order of the score $sc_{FIN}()$ described in Section 3.5 was different from that of the score of TSUB-AKI. Lingua::JA::Summarize::Extract could not extract the appropriate passages from

document sets on the basis of $sc_{FIN}()$, even though the document sets contained the same appropriate passages that were extracted by the proposed method. Therefore, summarization for a general purpose is insufficient for generation of mediatory summaries, and the proposed method can provide more appropriate mediatory summaries.

However, the results of the queries "Is catch and release a better way of fishing?" and "Does carbon dioxide cause global warming?" in Table 2 show that the proposed method could not extract all the appropriate passages that were extracted by Lin$_{TSU}$. We aim to improve the proposed method in future research.

5 Conclusion

We proposed a method for automated generation of mediatory summaries in order to facilitate users' assessment of the credibility of information on the Web. A mediatory summary is generated by extracting a passage from Web documents on the basis of their relevance to a given query, fairness, and density of keywords, which are features of the summaries constructed to determine the credibility of information on the Web. We demonstrated the effectiveness of the generated mediatory summary in comparison with the summaries of Web documents produced by Web search engines.

Acknowledgement

References

Juffinger, Andreas, Michael Granitzer, and Elisabeth Lex. 2009. Blog credibility ranking by exploiting verified content. In *Proceedings of the Second Workshop on Information Credibility on the Web (WICOW 2009)*, pages 51–57.

Kaneko, Koichi, Hideyuki Shibuki, Masahiro Nakano, Rintaro Miyazaki, Madoka Ishioroshi, and Tatsunori Mori. 2009. Mediatory summary

---

[7]We use precision, which represents #*correct outputs*/#*total outputs*, as an evaluation measure because it is difficult to calculate the recall of the mediatory summaries that are dynamically generated from Web documents and because users tend to read just a few of the summaries generated by the system.

generation: Summary-passage extraction for information credibility on the web. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC 23)*, pages 240–249.

Kawahara, Daisuke, Tetsuji Nakagawa, Takuya Kawada, Kentaro Inui, and Sadao Kurohashi. 2009. Summarizing evaluative information on the web for information credibility analysis. In *Proceedings of the 3rd International Universal Communication Symposium (IUCS 2009)*, pages 187–192.

Lin, Chin-Yew and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the Human Language Technology Conference 2003 (HLT-NAACL-2003)*, pages 71–78.

Miyazaki, Rintaro, Ryo Momose, Hideyuki Shibuki, and Tatsunori Mori. 2009. Using web page layout for extraction of sender names. In *Proceedings of the 3rd International Universal Communication Symposium (IUCS 2009)*, pages 181–186.

Murakami, Koji, Eric Nichols, Suguru Matsuyoshi, Asuka Sumida, Shouko Masuda, Kentaro Inui, and Yuji Matsumoto. 2009. Statement map: Assisting information credibility analysis by visualizing arguments. In *Proceedings of the Second Workshop on Information Credibility on the Web (WICOW 2009)*, pages 43–50.

Nakano, Masahiro, Hideyuki Shibuki, Rintaro Miyazaki, Madoka Ishioroshi, Koichi Kaneko, and Tatsunori Mori. 2010. Construction of text summarization corpus for the credibility of information from the web. In *Proceedings of the 7th Language Resources and Evaluation Conference (LREC 2010)*, pages 3125–3131.

Ohshima, Hiroaki, Satoshi Oyama, Hiroyuki Kondo, and Katsumi Tanaka. 2009. Web information credibility analysis by geographical social support. In *Proceedings of the 3rd International Universal Communication Symposium (IUCS 2009)*, pages 193–196.

Shinzato, Keiji, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto, and Sadao Kurohashi. 2008. Tsubaki: An open search engine inflastructure for developing new information access methodology. In *Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP 2008)*, pages 189–196.