

## INVESTIGATION ON EFFECT OF SNIPPET ON USER'S RELEVANCE JUDGMENT OF DOCUMENTS

YASUFUMI TAKAMA<sup>1</sup> MINGHUANG CHEN<sup>1</sup> SEIJI YAMADA<sup>2</sup>  
*Graduate School of System Design, Tokyo Metropolitan University*  
*National Institute of Informatics*

**ABSTRACT**—This paper investigates the effect of a snippet on users' relevance judgment of a document. Web search engines are becoming very useful tools in our daily life. One of the most important features of modern search engines is a snippet, which is expected to help users to find relevant pages immediately. Therefore, improving the quality of a snippet is important for minimizing the cost of user feedback required for finding relevant pages. This paper investigates the effect of a snippet on users' relevance judgment of a document by comparing users' relevance judgment when a snippet is provided and those without providing a snippet. The experimental results show providing snippets reduces users' judgment time, while keeping judgment accuracy. It is also shown the effect of snippet length on judgment time is strong when judging relevant documents. The obtained results will contribute to the improvement of a snippet generation method in terms of minimal user feedback.

Key Words: Minimal user feedback, snippet, relevance judgment, document retrieval

### 1. INTRODUCTION

When we are going to use a web search engine to find out the information that we need, we care about two things. One is the time cost, and the other is the relevance degree of the information we found in terms of our information need. A snippet is a fragment of a document, which usually represents its contents from the viewpoint of a query. It helps a user to easily make a judgment on whether or not to read the corresponding document.

Recently, the concept of minimal user feedback (MUF)[3] has been proposed, which aims at decreasing the cost of a user to find relevant information. As above-mentioned, a snippet is one of the important approaches for minimizing user feedback required for finding relevant Web pages. However, it has not yet been fully studied the effect of a snippet on users' relevance judgment, in terms of their judgment time and accuracy.

This paper investigates the effect of a snippet on users' relevance judgment by comparing users' relevance judgment when a snippet is provided and those without providing a snippet. A test interface is implemented, on which a user performs the task of judging a document. The experimental result shows providing snippets reduces users' judgment time, while keeping judgment accuracy. It is also shown the effect of snippet length on judgment time is strong when judging relevant documents.

### 2. RELATED WORKS

#### 2.1 Snippet Generation

There are a number of search engines on the Web, such as [www.google.com](http://www.google.com), [www.baidu.com](http://www.baidu.com) and [www.yahoo.com](http://www.yahoo.com) in recent year. Most of them have the similar interfaces for viewing the search results, which shows a title, URL and a snippet for each retrieved page. A snippet is an extracted fragment of a retrieved page, in which the keywords inputted as a query are usually highlighted. All of the presented information helps users to make decisions on whether or not to visit the corresponding page. Among them, a snippet is the most powerful clue for users to guess the content of the page.

In order to generate a snippet that is useful for users, it is an important matter how to pick out the fragments of a web page that reflects its contents. It is also important to generate query-oriented snippet, which reflects the part of a page that relates with a user's query.

The simple approach for generating a snippet is based on the technology of important sentence extraction in the field of text summarization. It calculates the score of sentences in a page according to a query,

and the sentences with the highest score are chosen as a snippet. This algorithm is further divided into various algorithms based on a method and attributes used for score calculation.

One of the simple methods for score calculation is based on the terms in a given query (query terms). As a query consists of terms that reflect a user's search purpose, a sentence that contains query terms is expected to refer to the topic of interest. Takami [5] has classified snippets into 4 types, for each of which appropriate term weighting scheme, such as TF, TF-DF, and TF-iDF, is proposed.

Li [2] has proposed a different algorithm from the above-mentioned score-based algorithms. It employs the language model to determine the segment of a document that is suitable as a snippet. Xue [6] has proposed image snippet, which extracts representative images from a page based on visual cue-based page segmentation and text-based similarity between a query and the segments.

## **2.2 User Behaviors Using Search Engine**

When a user obtains a search result from a search engine, s/he has to decide on which pages to visit. As noted in Sec. 2.1, a search result page contains various information that can be clues to determine pages to be visited. If we know the users' behaviors on viewing a search result page, we can improve the page design so that users can make their decisions with less effort.

From this viewpoint, users' behaviors of viewing web pages / search result pages have been studied by many researchers [1,4]. Recent trend is to track the users' eye movements using an eye-tracking device.

Rodden [4] has explored the relationship between mouse movements and eye movements when performing a search task with using Google. Various interesting patterns, such as keeping the mouse still while reading, and using the mouse as a reading aid are observed.

Cutrell [1] has analyzed the effect of various factors such as snippet length and the rank of the best result within a search result page on users' viewing behaviors. The experimental results show that the rank of the best result affects total time of performing a task, in particular for informational task. It is also shown that longer snippet increases total time but decreases click accuracy for performing a navigational task. On the other hand, longer snippet decreases total time but increases click accuracy for performing an informational task. Although this study investigates the effect of snippets on users' behaviors of examining search result pages, it does not directly discuss the contribution of snippet to relevance judgment of the content of a retrieved page.

## **3. EXPERIMENT ON EFFECTS OF SNIPPET ON USER'S RELEVANCE JUDGMENT**

### **3.1 Outline of Experiment**

An experiment is performed to compare users' relevance judgment when a snippet is provided and those without providing a snippet. The aim of the experiment is to investigate the effect of a snippet on relevance judgment of the corresponding document.

As for a target document set and queries, NTCIR-5 CLIR collection<sup>1</sup> is used. The collection consists of several document sets, among which Mainichi Daily News 2000-2001 (English-language newspaper of Japan's Mainichi Newspapers) is used in the experiment. For the document set, the following 3 topics are used in the experiment.

- #011: Find reports on Ichiro's first year in Major League after his move from a Japanese league.
- #037: Find documents describing how the percentage supporting the Mori Cabinet changed in 2001 as a result of Prime Minister Mori's handling of the Ehime-Maru accident.
- #041: Find documents describing the background to the case of large-scale food poisoning caused by Snow Brand dairy products.

For each query, the following number of documents is selected from the document set, for each of which a participant judges the relevance degree to the query with 4 levels: S (high relevant), A (relevant), B (partially relevant), and C (irrelevant).

---

<sup>1</sup> <http://research.nii.ac.jp/ntcir/permission/ntcir-5/perm-en-CLIR.html>

- #011: S=5, A=5, B=5, C=5 (Total=20 documents)
- #037: S=5, A=5, B=0, C=10 (Total=20 documents)
- #041: S=2, A=8, B=4, C=6 (Total=20 documents)

Figure 1 shows the screenshot of a test interface, which can be accessed from ordinary Web browsers. In the experiment of each topic, a brief description of the topic is displayed, and then the information about a document, either snippet or its original text, is displayed one by one. When a snippet is generated, terms in CONC field (specifying keywords relevant to whole topic) of the topic description are extracted, and the score of a sentence in a document is calculated based on the number of the terms appeared in it.

The presented order and displayed information (snippet / original text) of a document is randomly determined. Among the 20 documents, snippets and original text are displayed for 10 documents, respectively. When a snippet is displayed, a participant can optionally see its original text by clicking the button at the bottom of the page. A participant can also view the query description by clicking the button at the bottom of the page. When clicking those buttons, a Web browser displays the original text page and query description page, respectively.

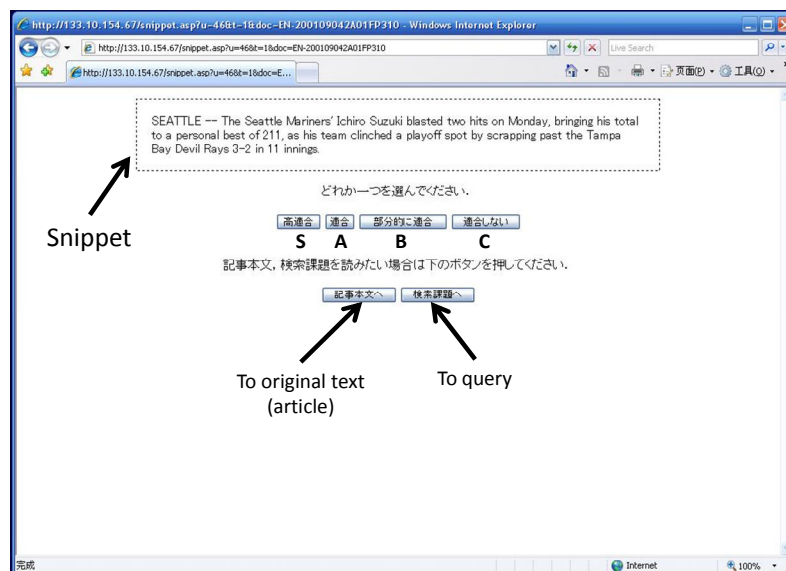


Figure 1 Screenshot of Test Interface

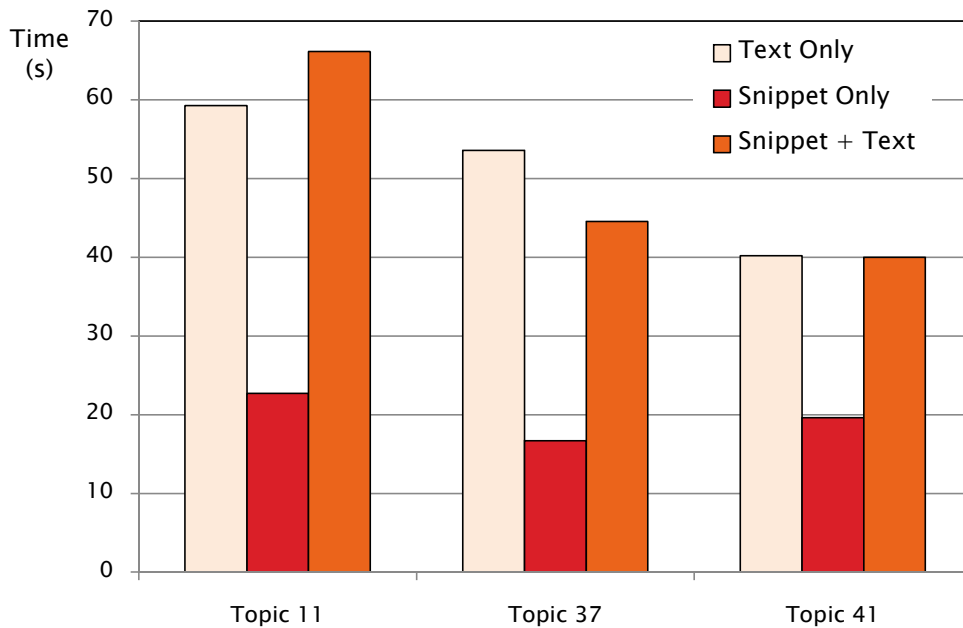
#### 4. EXPERIMENTAL RESULTS

Total number of participants is 14, which includes 6 undergraduate students, 7 graduate students, and 1 researcher. We asked every participant to take part in the experiment for 3 topics. However, some participants answered 1 or 2 of those topics. The number of participants was 13 for topic #011, 11 for #037, and 12 for #041, respectively.

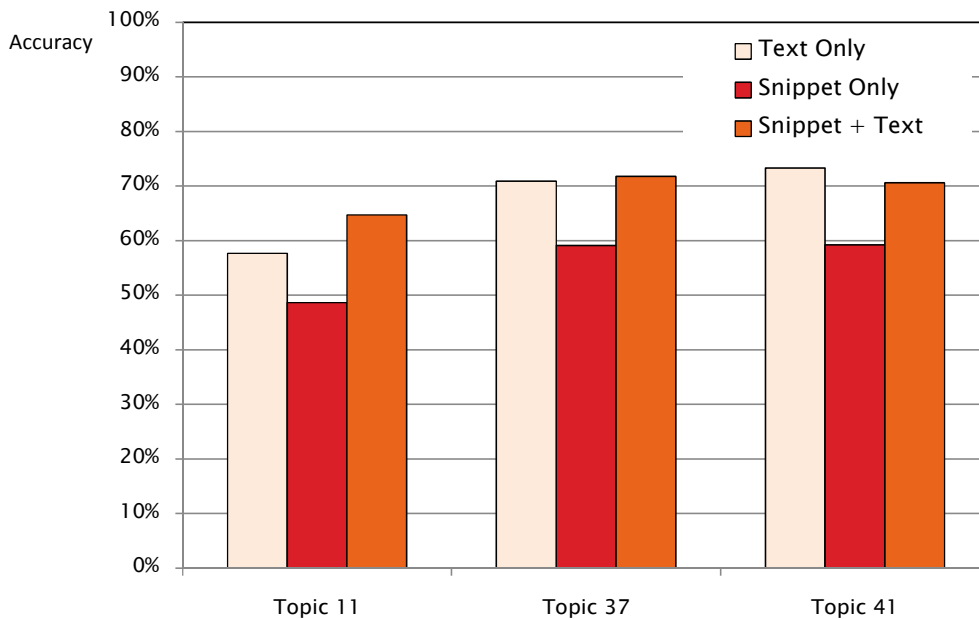
For each topic, a participant judged 10 documents based on snippets, and other 10 documents based on original text. Therefore, total 360 (36x10) data of judging documents based on original text was obtained.

On the other hand, when a document is presented with a snippet, a participant can optionally read its original text as noted in the previous section. Among 360 data of judging documents based on snippets, 287 data (79.7%) correspond to the case when a participant judged the document with a snippet only, and the remaining 73 data correspond to the case when judged with both a snippet and its original text. Therefore, in the rest of this section, the results are divided into the following 3 cases.

- Snippet Only: a snippet is presented and participants made decision without reading original text.
- Snippet+Text: a snippet is presented and participants also read its original text.
- Text Only: original text is presented.



**Figure 2 Comparison of Judgment Time among Conditions**



**Figure 3 Comparison of Accuracy (Rigid) among Conditions**

Figure 2 shows the comparison of judgment time among 3 conditions. It clearly shows judgment time when using snippet only is much shorter than other conditions.

Figure 3 and 4 show the comparison of judgment accuracy (precision) among 3 conditions. As the NTCIR-5 CLIR test collection has 2 types of criteria, rigid and relaxed, the results are shown with both criteria. Rigid precision (Fig. 3) is calculated when S and A are regarded as relevant, whereas relaxed precision (Fig. 4) is calculated when S, A, and B are regarded as relevant. Comparing rigid and relaxed criteria,

precision under rigid criterion is less than 80% even when participants judged based on original text. Therefore, we think relaxed criterion is more realistic than rigid as an evaluation criterion. In the relaxed criterion, all of three conditions achieve more than 80% precision, which is supposed to be acceptable results.

Although the precision of the snippet only condition is lower than other 2 condition for both cases, there is not so big difference. That is, in the rigid condition, the biggest difference between text only and snippet only is 14.1 point (topic 41), which means the difference of less than 2 documents per a participant judging 10 documents. In the case of relaxed condition, such differences are less than 1 document for all of the 3 topics. Therefore, it can be seen from Fig. 2-4 that when participants read only snippets, the judgment time is reduced while the degradation of precision is small compared with original text is presented.

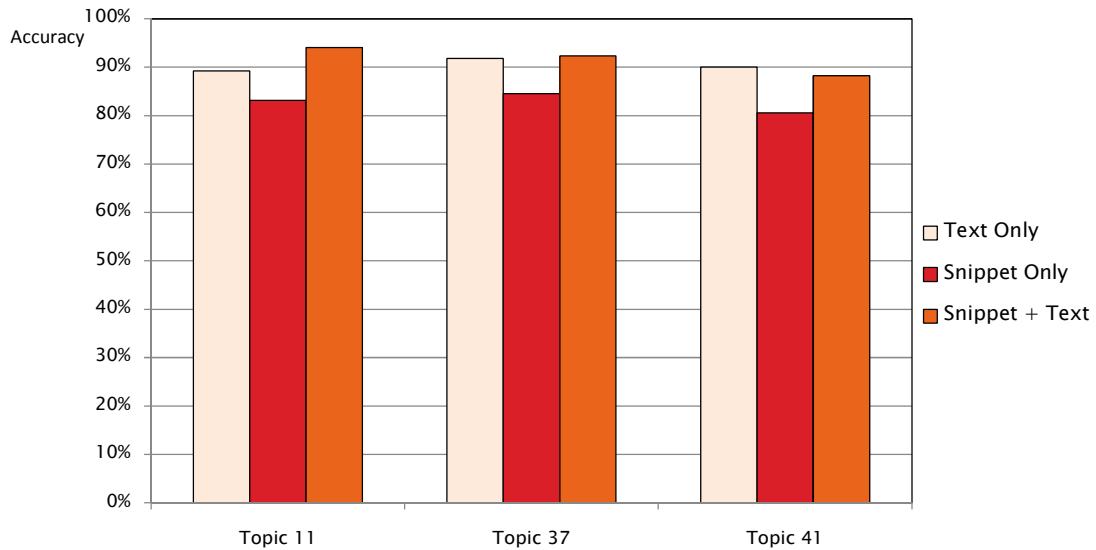


Figure 4 Comparison of Accuracy (Relaxed) among Conditions

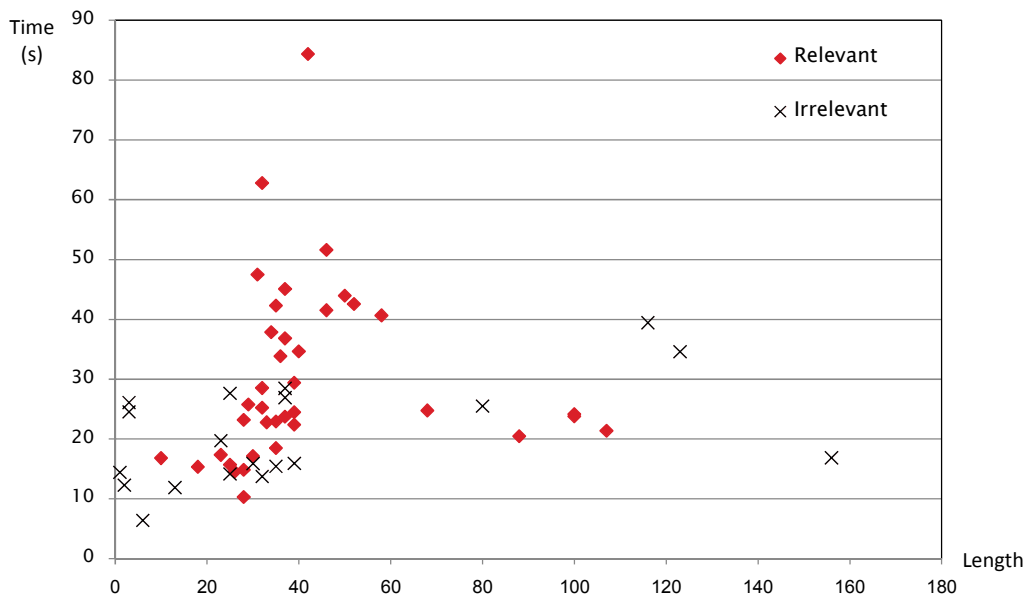


Figure 5 Scatter Plot of Snippet Length versus Judgment Time

We think the reason why a snippet can reduce users' judgment time while keeping judgment accuracy is because it can provide only the information enough to grasp the contents of a document. When the length of a snippet increases, the possibility of containing relevant information will increase. On the other hand, the possibility of containing unnecessary information will also increase, which will increase the judgment time. Therefore, the length of a snippet is one of the important factors regarding its quality.

In order to examine the effect of snippet length on the relevance judgment, the relationship between snippet length and judgment time is examined. Fig. 5 shows the scatter plot of snippet length versus judgment time. In the figure, different symbols are used for relevant (S, A, B) and irrelevant (C) documents. From the figure, it can be seen that there is different tendency between relevant and irrelevant documents. In the case of relevant documents, judgment time is drastically increased when snippet length is less than 60. On the other hand, in the case of irrelevant document, judgment time is also proportional to snippet length, but its slope is gentle compared with relevant documents. It is also noted that we cannot find such clear relationship between snippet length and judgment accuracy.

We suppose that the reason why snippet length has stronger relation with relevant documents than irrelevant documents is because user's judgment process consists of two stage. First, a user perform binary judgment, i.e. a document is judged as either relevant or irrelevant. When a document is judged as relevant, s/he will then estimate its relevance degree, i.e. either S, A, or B as the second phase. Therefore, judging relevant documents takes more time than judging irrelevant documents. In order to verify this assumption, further experiments will be required, which is one of our future studies.

## 5. CONCLUSIONS

In this paper, the effect of a snippet on users' relevance judgment is investigated. The experiment is performed to compare users' relevance judgment when a snippet is provided and those without providing a snippet. The experimental results show providing snippets reduce users' judgment time, while keeping judgment accuracy. It is also shown the effect of snippet length on judgment time is strong when judging relevant documents. Future works include the collection of more experimental results from many participants as well as more detailed analysis of the results using eye tracking device. Based on the analysis of the results, we also plan to improve the quality of a snippet from the viewpoint of minimizing user feedback.

## REFERENCES

1. E. Cutrell and Z. Guan: An eye-tracking study of information usage in Web search: Variations in target position and contextual snippet length, CHI'07, pp. 407-416 (2007).
2. Q. Li, K. S. Candan, Y. Qi: Extracting Relevant Snippets from Web Documents through Language Model based Text Segmentation, WI-IAT2007, pp. 287-290 (2007)
3. M. Okabe and S. Yamada: Semi-supervised Query Expansion with Minimal Feedback, IEEE Trans. Knowledge and Data Engineering, Vol.19, No.11, pp.1585-1589 (2007)
4. K. Rodden and X. Fu: Exploring How Mouse Movements Relate to Eye Movements on Web Search Results Pages, SIGIR 2007 Workshop on Web Information Seeking and Interaction (WISI), pp.29-32 (2007)
5. S. Takami and K. Tanaka: Customizing Web Search Results by Dynamic Re-Generation of Web-Snippet based on Search Purpose, DBSJ Letters, Vol. 6, No. 2, pp. 33-36 (2007) (in Japanese)
6. X. Xue, Z. Zhou, Z. Zhang: Improve Web Search Using Image Snippets, Proc. 21st National Conference on Artificial Intelligence, pp. 1431-1436 (2006)