

Sentence length, sentence fragment and images affecting presentation of search result pages

Hong Xu^{1,2,a*}, Ye Jiang^{2,b*}, Yi-min Wang^{3,c}, Ye-wei Sun^{4,d} and Xue-qing Li^{1,e}

¹Shandong University, China

²Shandong Institute of Commerce and Technology, China

³University of Jinan, China

⁴Virginia Polytechnic Institute and State University, America

^axuhonglele@163.com, ^bzgjiangtoby@163.com, ^cymking6688@163.com, ^dyeweis12@vt.edu, ^exqli@sdu.edu.cn

Keywords: search result; presentation style; presentation feature; click-through behaviors.

Abstract. Presentation styles of search results have been constantly changing in these years, which affect users' click-through behaviors. In order to discover what the efficient presentation styles are, in this study, we investigate three features of search results: sentence length, full sentence/ sentence fragment and with/without image. These three features were compounded into twelve combinations and tested by 41 participants in our study. It is found that the most efficient search result presentation style should be formed by long sentence length, full sentence and with image, followed by short sentence length, full sentence, with image. The badly designed presentation style is without image, full sentence for all kinds of sentence lengths. The longer the sentence length, the better the result. When without image, sentence fragment has better performance than full sentence. Full sentence is more efficient than sentence fragment with image for short and long sentence lengths, but not for medium one.

Introduction and related work

There are numerous presentation features in search result pages, which strongly affect searchers' click-through behaviors [1, 2]. A well-designed search result presentation should be informative and time-saving. A study indicated that a participant took 7.78 seconds to select a relevant search result averagely, and 11 seconds for some difficult queries [5]. Another research found participants spent 85% of the total search time on the first result, but only 7.5% of the time on the second [12]. Participants rarely looked at the second page if they couldn't find the related information on the first page. They chose to reformulate the query terms by using different searching keywords, or try to use other search strategies [3, 4]. In order to determine which of the many presentations of search result leads to substantial information for search engines result page, this research aims to compare different combinations of presentation features and identify which of them are the informative and time-saving.

In this study, three features are chosen by establish a survey of asking participants' subjective opinions on the search result page. The outcome indicated the three features, which are sentence length, sentence fragment and with/without image, are most concerned during their search processes. Sentence length and sentence fragment are shown by the snippet, which below the search result title. The search result snippet contains phrase of search query in the search result page. Although long sentences provide more information than short one, the accuracy of relevant result is critical for participants. A study divided the sentence length into three levels in the snippets: short (1 line), medium (2-3 lines) and long (over 4 lines). By applying an eye tracking experiment, they found participants preferred to long sentence length because it shows more information than short one [6]. However, a study identified that the relevant sentences shown in the result page were simply chosen by the longest sentence in the

systems of the TREC Novelty track data. So the real relevant sentence might be ignored when the system retrieving [14].

Another issue is whether to show sentence fragments or full sentences in the search result snippets. Nowadays, search engines are using query-biased snippet, which is formed by sentence fragments, to extracts query terms from the document and presents highlighted query terms in the snippets. The information in the snippet can be simply and accurately understood when the integrated query terms are highlighted. However, a research indicated that searchers feel difficult to identify the relevant results when the query terms are too generic. Moreover, sentence fragments are usually cut-off illogically and hard to understand. Another research also found sentence fragment is ambiguous when the search query is complex [13]. On the contrary, the full sentence enlarges the text field and provides more information in the snippets [15].

The blended search result page interface is applied by main-stream search engines, such as Bing, Google etc. A study found the searchers prefer results with image presentation in the result page [1]. Some researchers also found searchers were more successful at finding relevant result with images [16, 17, 18]. In this study, we analyzed whether to combine search results with/without image.

Methodology

We conducted a demographic questionnaire to collect the background of participants. In the last section of this questionnaire, we asking participants what they think presentation features of a search result page should be considered when designing the search result page, and the chosen features of search result presentation in this study were based on that question. 41 participants were answered this questionnaire. 76% answered image, 13% answered sentence length and 11% answered sentence fragment. The result turns out that the features are sentence length, sentence fragment/full sentence in a snippet and with/without image. Although these features were analyzed individually, it is not known yet how these features would affect participants’ click-through behavior if they are combined in various ways. We established an online experiment, which is a simulation of search engine result pages similar to Google’s search result pages, to integrate these three features into twelve combinations (shown in Table I) to capture participants’ preference. We built these pages by using Wix.com, which is a cloud-based website platform, to tests these three features in these pages. In these pages, participants went through these combinations within the four tasks and complete a survey of relevance judgment at the bottom of that page.

Table 1. The twelve combinations of three features

| Combinations | Sentence length | Sentence fragments | Image |
|--------------|-----------------|--------------------|-----------|
| 1 | Short | Full sentence | Include |
| 2 | Short | Full sentence | Non image |
| 3 | Short | Sentence fragment | Include |
| 4 | Short | Sentence fragment | Non image |
| 5 | Medium | Full sentence | Include |
| 6 | Medium | Full sentence | Non image |
| 7 | Medium | Sentence fragment | Include |
| 8 | Medium | Sentence fragment | Non image |
| 9 | Long | Full sentence | Include |
| 10 | Long | Full sentence | Non image |
| 11 | Long | Sentence fragment | Include |
| 12 | Long | Sentence fragment | Non image |

Each task included a topic, which is based on the TERC (2005) document collection (as Figure 1), to introduce a subject. Firstly, we built an introduction page and gave objectives of each task before participants start to browse search result pages. Secondly, a short description would explain what relevant documents should be and a narrative gave details of the information relate to that topic. So participants could know the criteria of how to complete these tasks successfully in the introduction page. TERC (2005) also gave the relevance level of each document. The topics and documents are randomly chosen from TERC, the relevant level ranges from 0 to 2 (0 represent Non-relevant, 2 represent highly relevant). TERC also gives the description and narrative of each topic, which help participants to make the judgment of relevance level. Afterwards, they went through the search result pages that contained nine results respectively. Finally, they decided relevance level and their outcomes were recorded by a survey at bottom of each page.

| 1 | Document ID | Task | TERC Relevance Level |
|----|------------------|-------|----------------------|
| 2 | APW19990615.0145 | 310-1 | 2 |
| 3 | APW19980603.0791 | 310-1 | 0 |
| 4 | XIE20000920.0016 | 310-1 | 1 |
| 5 | APW19980616.0702 | 310-1 | 0 |
| 6 | APW19980625.0122 | 310-1 | 1 |
| 7 | APW19980627.0596 | 310-1 | 0 |
| 8 | APW19990724.0024 | 310-1 | 2 |
| 9 | APW19980702.0863 | 310-1 | 0 |
| 10 | APW20000511.0135 | 310-1 | 1 |
| 11 | NYT19990405.0050 | 310-2 | 0 |
| 12 | NYT19990405.0085 | 310-2 | 0 |
| 13 | NYT19990524.0289 | 310-2 | 1 |

Figure 1: The sample of the TERC relevance level of each specific document and task.

1. Sentence length

In terms of sentence length, in order to optimize the number of words shown in a snippet, different sentence length could affect participants' click-through differently. Many researchers found that there is a tradeoff between displaying a long sentence and a short sentence in the search snippets [6,7,8]. To be specific, a large search result snippet contains approximately over 4 lines of sentences, the information could be easily identified by participants by just looking at the snippets rather than clicking into that page. However, participants take more time on reading these long sentences. Long sentence might also cause intensive layout in the search result page so as to reduce the readability. In order to analyze how many words should be displayed in a snippet, sentence length has been divided into three levels as the following:

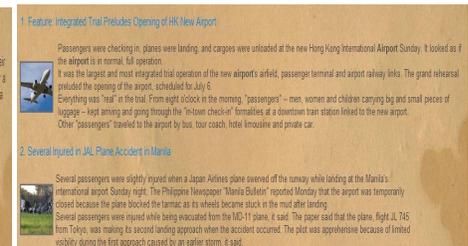
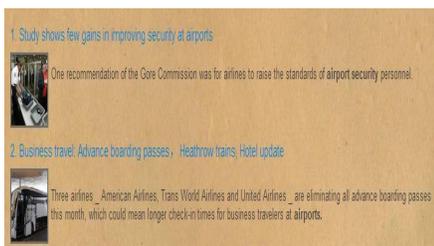


Figure 2. Short sentence length Figure 3. Medium sentence length Figure 4. Long sentence length

2. Sentence fragment

It is quite often to use sentence fragments in web search engines currently, because its precision and simpleness. It helps to improve the readability of snippets, but participants might be misled due to these incomplete sentences. [9] found that sentence fragments have negative effects, because the meaning of these sentence fragments are ambiguous and illogically. Moreover, [10] conducted an experiment and found that participants prefer full sentences rather than sentence fragments. In order to find out the

optimized combination of our three features, sentences are categorized into two patterns: sentence fragments and full sentences as shown below:

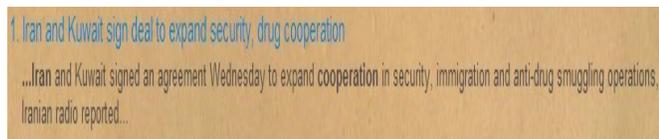


Figure 5. Sentence fragment

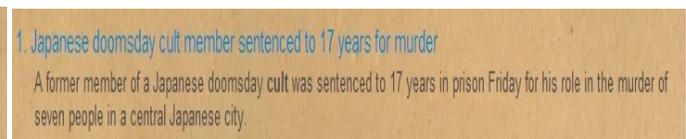


Figure 6. Full sentence

3. Image included or non-image

Commonly, many search engines have embedded multimedia resources in their search engine result pages to improve readability for the participants [11]. [1] found that the factors of blended search interface would affect participants' click-through behavior. As they found that the different source types are important factors for searchers to determine the relevance in the blended search result pages. The graphical element would attract participants' attention when they look at the search result page at the first time. In this study, we also blend image resources in the search engine results, In order to find out the optimized combination of our three features, with or without image is combined with various sentence lengths, sentence fragments or full sentences and are evaluated.

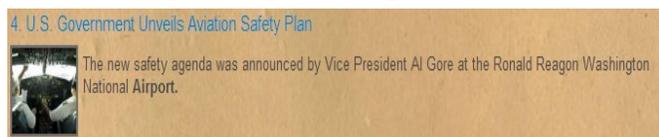


Figure 7. Search result with image

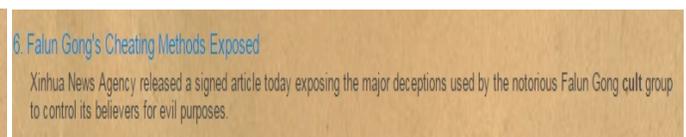


Figure 8. Search result without image

Data collection

1. Assumptions

Two types of data were collected for this study: the initial and secondary data. The initial data represent participants' relevance judgment of each search result. Several independent surveys allowed participants to fill in their judgment at the bottom of each result page. They could click into each search result page to have a further reading then make their decisions about relevance. The initial data was collected by SurveyMonkey, which is a free web survey development cloud based tool for creating online survey. Our first assumption is: if the percentage of correct relevance level is high, the combination is explicit to guides participants to find the correct results.

The secondary data was the numbers of page views obtained by the tool: Google Analytics. Our second assumption is: if a particular page has a large number of views, it means the participants may feel confused to determine the page's relevance. On the contrary, if a result page received less number of views, this means its relevance level can be easily determined by the participants.

| Page | Pageviews | Unique Pageviews | Avg. Time on Page | Entrances | Bounce Rate | % Exit | Page Value |
|---|---|---|---|---|---|---|---|
| | 919 % of Total: 100.00% (919) | 792 % of Total: 100.00% (792) | 00:00:52 Site Avg: 00:00:52 (0.00%) | 126 % of Total: 100.00% (126) | 26.19% Site Avg: 26.19% (0.00%) | 13.71% Site Avg: 13.71% (0.00%) | \$0.00 % of Total: 0.00% (\$0.00) |
| 1. /http://www.yejiang18.com/#task-341/cf18 | 108 | 100 | 00:01:13 | 0 | 0.00% | 7.41% | \$0.00 |
| 2. /http://www.yejiang18.com/ | 102 | 97 | 00:00:44 | 91 | 35.16% | 36.27% | \$0.00 |
| 3. /html/56643b_ec47f077e896674841a6e1982c9_0962d.html | 85 | 27 | 00:00:24 | 27 | 0.00% | 31.76% | \$0.00 |
| 4. /http://www.yejiang18.com/#introductory-questionsnaire/c18wx | 68 | 62 | 00:01:20 | 1 | 0.00% | 26.47% | \$0.00 |
| 5. /http://www.yejiang18.com/#task-310/c21k | 58 | 49 | 00:00:09 | 0 | 0.00% | 1.72% | \$0.00 |
| 6. /http://www.yejiang18.com/#task-instruction/cv65 | 42 | 42 | 00:00:54 | 0 | 0.00% | 16.67% | \$0.00 |

Figure 9. The sample of page views output by Google Analytics

As the two types of data collected, the informative presentation of combination should be assumed as: less number of views of that particular page and the most of participants matched with the correct relevance level from TERC.

2. Initial data--The number of participants matched TERC relevance

Table 2 shows the percentage of participants discovered the correct relevance level in the task 341. Task 341-2 has the lowest percentage of participants discovered the correct relevance level with 41.2%. When we decrease the sentence length from medium to short, the percentage of participants discovered the correct relevance level is increase to 55.7%. The task 341-1 consists of short sentence length, full sentence and image, and it is the second highest of percentage of participants discovered the correct relevance level. However, the highest percentage of participants discovered the correct relevance level is task 341-3, the sentence length is long, and it includes full sentence and image feature.

Table 2. The participants' relevance judgement from Task 341-1 to Task 341-3. %Discover= the percentage of participants discovered the correct relevance level. %Mistake= the percentage of participants chose the incorrect relevance level.

| Task | Combination | % Discover | % Mistake |
|-------------|--|-------------------|------------------|
| 341-1 | Short sentence length, Full sentence, Image | 55.70% | 45.30% |
| 341-2 | Medium sentence length, Full sentence, Image | 41.20% | 59.80% |
| 341-3 | Long sentence length, Full sentence, Image | 67.30% | 33.70% |

In the task 310, when the full sentence converts to the sentence fragment, the percentage of the percentage of participants discovered the correct relevance level averagely decreases by around 10%. The task 310-1 is the combination of short sentence length, sentence fragment and image, and it has the lowest percentage in the task 310. We increase the sentence length to the medium in the task 310-2, and the percentage is interestingly increased from 43% to 45.9%. In the task 310-3, when the sentence length is long, the percentage is also increase to 50.1%. The sentence length is the same as task 341, which we increase its length from short to long, the sentence fragment is the main factor that decreased the percentage of participants discovered the correct relevance level.

Table 3. The percentage of participants discovered the correct relevance level from Task 310-1 to Task 310-3. %Discover= the percentage of participants discovered the correct relevance level. %Mistake= the percentage of participants chose the incorrect relevance level

| Task | Combination | % Discover | % Mistake |
|-------------|--|-------------------|------------------|
| 310-1 | Short sentence length, Sentence fragment, Image | 43.00% | 57% |
| 310-2 | Medium sentence length, Sentence fragment, Image | 45.90% | 55.00% |
| 310-3 | Long sentence length, Sentence fragment, Image | 50.10% | 50.00% |

Table 4 depicts the percentage of participants discovered the correct relevance level in Task 325. It is obvious that the percentage is lower than others three tasks. Specifically, Task 325-1 consists of short sentence length, full sentence and non-image. The percentage of participants discovered the correct relevance level is the lowest in the task 325. However, the percentage of participants discovered the correct relevance level is growing when we increase the sentence length from short to medium. The task

325-2 shows that 37.8% of participants found the correct relevance level. The highest percentage of participants discovered the correct relevance level is 43.5% in the task 325-3.

In general, the image could be an essential element that attracts participants’ attention, but we remove image feature in the task 325, so the percentage of participants discovered the correct relevance level is relatively lower than other tasks. When there is no image included in the search results, the snippets are the only features could guide the participants’ judgment, so the longer sentence length in the snippets the more information is presented to the participants.

Table 4. The percentage of participants discovered the correct relevance level from Task 325-1 to Task 325-3. %Discover= the percentage of participants discovered the correct relevance level. %Mistake= the percentage of participants chose the incorrect relevance level

| Task | Combination | % Discover | % Mistake |
|-------------|---|-------------------|------------------|
| 325-1 | Short sentence length,Full sentence,No image | 34.90% | 66.00% |
| 325-2 | Medium sentence length,Full sentence,No image | 37.80% | 63.00% |
| 325-3 | Long sentence length,Full sentence,No image | 43.50% | 57.00% |

The task 330 change the full sentence to sentence fragment so as to find out whether the sentence fragment could improve participants’ relevance judgment in the non-image snippet. The Table 5 shows the percentage of participants discovered the correct relevance level is only 40% in the task 330-1. When we increase the sentence length from short to medium, the percentage of participants discovered the correct relevance level is also growing from 40% to 44%. The highest percentage of participants discovered the correct relevance level is 47% in the task 330-3.

It is easy to find that the sentence length is an important feature when the search results exclude the image. The long sentence length performs better than medium and short sentence. Full sentence displayed in the search result summaries has positive effect for the presentation as well. In the task 330, the percentage of participants discovered the correct relevance level is averagely higher than the task 325

Table 5. The percentage of participants discovered the correct relevance level from Task 330-1 to Task 330-3. %Discover= the percentage of participants discovered the correct relevance level. %Mistake= the percentage of participants chose the incorrect relevance level

| Task | Combination | % Discover | % Mistake |
|-------------|---|-------------------|------------------|
| 330-1 | Short sentence length,Sentence fragment,No image | 40.00% | 60.00% |
| 330-2 | Medium sentence length,Sentence fragment,No image | 44.00% | 56.00% |
| 330-3 | Long sentence length,Sentence fragment,No image | 47.00% | 53.00% |

3. Secondary data—Page views

If a specific page received a large number of page views, it means that page has unclear presentation style and it slows down participants’ relevance judgment. On the contrary, if a page received less number of page views, this page has clear presentation style and it is easily to be identified by participants.

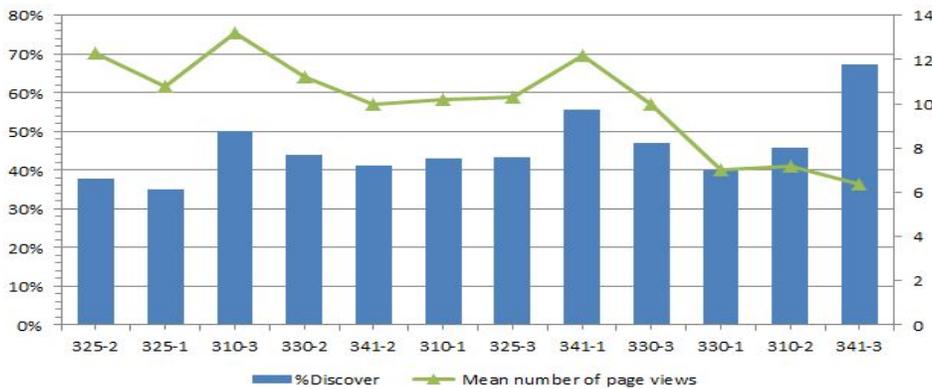


Figure 10. The mean of page views and %Discover in each task.

In this study, we define three levels to describe the performance of each combination, which are high, medium and low, for the number of page views. If the numbers of page views are equal to or less than 8, this means its corresponding combination has the best performance on presenting information, and it is the high level. If the numbers are between 8 and 10, this means the performance is medium. If the numbers are equal to or over 10, this means the performance is low.

As shown in the Fig.10, the task 341-3 has the lowest number of page views 6.4, which means this combination (Long sentence length, Full sentence and Image) could improve the presentation style. If the initial data of task 341-3 reached a high that means the participants can easily identify the relevance of these documents without click into each specific page repeatedly and this combination has the best performance of presenting the informative search result presentation. Otherwise, this combination confuses the participants' relevance judgment so they refused or omitted click into these pages. This blue part is not clear. On the other hand, Task 310-3 has the highest number of page views. It means that this combination may give ambiguous informative presentation of search results, so the participants have to click into that page once or at least once.

Discussion and Conclusion

1. The overall result analysis

Combining the initial and secondary data and do the comparison, we find that the task 341-3 has the highest % Discover value and the lowest number of page views throughout this study. Therefore, the task 341-3 is the most informative and efficient presentation style so that the participants could find the correct relevance level for each search result more efficiently. Followed by the task 341-1, it has the second highest % Discover value, however, its number of page views ranks at the second highest, thus its presentation style is ambiguous. On the contrary, the task 325-1 obtains the lowest % Discover value and a high number of page views, so its presentation style is not recommended.

2. Sentence length

All these fore tasks demonstrate that, compared with short and medium sentence length, the long sentence length gets the highest % Discover value. 75% of these fore tasks show that the medium sentence length obtain a higher % Discover value than the short ones. The result indicates that participants prefer to longer sentence length because it gives more sufficient information so that the participants could find the relevant sentence from the search result snippets instead of clicking into it.

3. Full sentence or sentence fragment

If we compare the task 325 and 330, i.e. when there is no image in the snippets, we could find out that the task 325 always has lower % Discover value and higher number of page views for each sentence length condition. Therefore, when there is no image contained in the snippet, sentence fragment is preferable to full sentence.

When compare the task 310-1 with 341-1, i.e. when there are small sentences and with image, we find out that full sentence works better than sentence fragment. 310-2 works much better than 341-2, which is to say, when there is medium sentence length and image included, sentence fragment has better performance than the full sentence. One should notice that when there is long sentence length and image contained, sentence fragment has remarkably bad performance, which has a relatively low % Discover value and the highest number of page views.

Although the sentence fragment could save a lot of space for the search result snippets, the participants may be disturbed by the illogical sentence pieces. However, both the long sentence length and full sentence have the limitation that they require more space in the search result page, which will reduce the number of relevant snippet shown on search result pages.

4. With/ without images

Comparing the task 325 (330) with 310 and 341, the result reveals that 92% snippets with image get higher % Discover value than that without image. There is not obvious result from the aspect of number of page views comparing with/without image conditions.

To conclude, we find the well-designed combination of search result features is: long sentence length, full sentence, with image, followed by short sentence length, full sentence, with image. The badly-designed combination of search result features is: full sentence, without image, regardless what the sentence length is.

The result indicates that the longer the sentence length of search result summaries, the better. The sentence fragment is more preferable to full sentence when the snippet does not contain an image. When there is an image in the snippet, full sentence has better performance than sentence fragment for short and long sentence lengths, and is on the contrary for medium sentence length.

References

- [1] Sushmita, S., Joho, H., Lalmas, M., & Villa, R. (2010, October). Factors affecting click-through behavior in aggregated search interfaces. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 519-528). ACM.
- [2] Chen, M., Hearst, M., Hong, J., & Lin, J. (1999). CHA-CHA: A SYSTEM FOR ORGANIZING INTRANET SEARCH RESULTS.
- [3] Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-Tracking Analysis of User Behavior in WWW Search.
- [4] Silverstein C, Marais H, Henzinger M, et al. Analysis of a very large web search engine query log[C]//ACM SIGIR Forum. ACM, 1999, 33(1): 6-12.
- [5] Hearst, M. (2009). Search user interfaces. Cambridge University Press.

- [6] Cutrell, E., & Guan, Z. (2007, April). What are you looking for?: an eye-tracking study of information usage in web search. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 407-416). ACM.
- [7] Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., & Karger, D. R. (2003, September). What makes a good answer? The role of context in question answering. In Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003) (pp. 25-32).
- [8] Kaisser, M., Hearst, M. A., & Lowe, J. B. (2008). Improving Search Results Quality by Customizing Summary Lengths. In ACL (pp. 701-709).
- [9] Rose, D. E., Orr, D., & Kantamneni, R. G. P. (2007, May). Summary attributes and perceived search quality. In Proceedings of the 16th international conference on World Wide Web (pp. 1201-1202). ACM.
- [10] White, R. W., Bilenko, M., & Cucerzan, S. (2007, July). Studying the use of popular destinations to enhance web search interaction. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 159-166). ACM.
- [11] iProspect. iProspect Blended Search Results Study. iProspect, April 2008. http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm.
- [12] Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology*, 56(6), 559-570.
- [13] Bando L L, Scholer F, Turpin A. Sentence length bias in TREC novelty track judgements[C]//Proceedings of the Seventeenth Australasian Document Computing Symposium. ACM, 2012: 55-61.
- [14] Bever T G, Garrett M F, Hurtig R. The interaction of perceptual processes and ambiguous sentences[J]. *Memory & Cognition*, 1973, 1(3): 277-286.
- [15] Alonso O, Baeza-Yates R, Gertz M. Effectiveness of temporal snippets[C]//WSSP Workshop at the World Wide Web Conference—WWW. 2009, 9.
- [16] Kules B, Capra R, Banta M, et al. What do exploratory searchers look at in a faceted search interface?[C]//Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. ACM, 2009: 313-322.
- [17] Hearst, M., Elliot, A., English, J., Sinha, R., Swearingen, K., & Yee, P. (2002). Finding the flow in web site search *Communications of the ACM*, 45(9), 42-49.
- [18] Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. In Proceedings of the SIGCHI Conference on Human factors in Computing Systems, Ft. Lauderdale, FL (pp. 401-408). New York: ACM Press.