1 Introduction

• OKSAT submitted 18 runs for NTCIR-14 OpenLiveQ-2 task.
• We reorder questions by using white and black words because most questions in OpenLiveQ-2 question-data of this task are fit for the queries.
• The white words are selected by the frequency in questions, Google suggest and/or manual.
• On the other hand, the black words are selected by the rarerness in the questions.
• The reorder of questions by white words is more effective than by black words from the evaluation results.

1 White and Black Words

• The question data (OpenLiveQ2-question-data.txt) has provided almost matching questions for each query.
• As we have ascertained in the runs of section 3, the Q-measure does not change much even if the top 10 questions for each query are moved to the bottom (runs 5-15) or all questions are sorted in the reverse order (runs 16-20).
• Therefore, we defined nouns that appear in questions and seem to fit to a query as white words. Conversely we defined nouns that seems difficult to fit to a query as black words.
• By moving the questions containing white words forward in the ranked list and moving the questions containing black words backward in the list, we considered that the questions fitted to the query can be gathered around top of the list.
• We set nouns which appear less frequently in the questions for each query as black words. On the other hand, we tried to set nouns as white words. (2) which appear many times in questions, (3) which are suggested by Google suggest, (4) which we found by manual, and (5) which are found in Wikipedia.

2-1 White and Black Words

Cn't'd

Fig. 1. Moving query by white and black

2-1 Submitted Runs

We submitted the following 18 runs. The numbers in parentheses are the original id's of the task.

• run-N3(114): A white word based run, using the Google Suggest API to extract related terms and sort in descending order of the number of occurrences in the questions.
• run-N4(119): A black word based run, with questions including proper nouns appearing only once in all questions for the query, followed by the rest in the initial order.
• run-N5(142): Minor change version of run-N4. Run that dynamically adjusted the number of occurrences of black words so that more than half of original TOP10 questions contain black words.
• run-N6(146): After separating the blacklist with run-N4 equivalent, the run that applied the sort of white word corresponding to run-N3 to non-blacklist.
• run-N7(153): Minor change version of run-N6. Run was evaluated by adding the white word including rate and the number of page view normalized by their maximum values.

2-3 Submitted Runs

Cn't'd

run-S3(132): The questions which have the plural black words (rare) in nouns in the questions for each query were ranked low in the ranked list. We extracted nouns by morphological analysis [45] of the title.
run-S4(151): Same ranking method were used in run-S3, however we changed the standard of rarerness (####).

run-U1(38): The same ranking method as run-9 of OKSAT [6] in OpenLiveQ-3. Questions were ranked by page view (square root of body length).
run-U2(135): The same ranking method as run-20 of OKSAT in OpenLiveQ-1.

2-4 Submitted Runs

Cn't'd

run-U0(94): We moved the top 10 questions behind the bottom.
run-U1(96): We sorted questions in descending order by the number of page view in questions for each query.
run-U2(98): We sorted the order of the original questions in reverse order.
run-U3(104): We sorted by score according to the number of page view and Q-measure. (The best run of our group.)
run-U4(116): Run-based high in the title of the question that contained many white words. We manually selected white words from the title of top-50 questions. The number of target queries was 101, among 1,000 queries in total.
run-U6(140): We ranked high in the title of the question that contained many black words. We manually selected black words from the title of top 13 questions. The number of target queries was 101.