Good Morning!

I am Hiroki Tanioka.

I am here because I got a notice from OpenLiveQ-2.

You can call me just "Hiroki".

NTCIR-14 OpenLiveQ-2

WHAT IS TARGET?

OpenLiveQ-2 requires sorted QA lists for each query to participants.

The queries are short queries which are composed of some keywords.

The QAs are released with some statistics including click through rate, views count, updated time, etc.

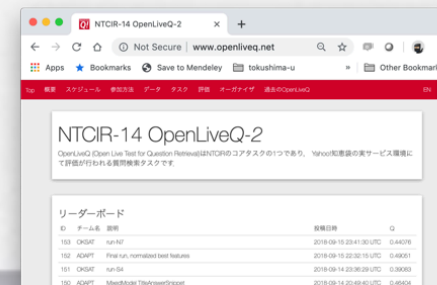1,000 QA list are for train, other 1,000 QA list are for test.

More info at http://www.openliveq.net/

HOW TO EVALUATE?

To evaluate submitted QA list, this task has two phases; offline test and online test.

Offline test means calculating accuracy in some mesures, nDCG, Q-measure, etc. using prepared answer list.

Online test means comparing superiority of submitted QA lists at Yahoo! Chiebukuro by live users.

"

Questions are commonly expressed so as to elicit information and to require resolution or discussion from users.

But, the readers are not the same.

What is catchy question?

To find out the statistics of catchy in QA systems, participated in OpenLiveQ-2.

I.
Offline Test

Why climb a mountain?
Because there is a mountain.

My Strategy to Climb the Mountain

- ▨ Research the last case in NTCIR-13
- ▨ Gathering available information
- ▨ Tuning based on like linear programming

Anyway, I climbed to the top of the mountain.

Let's review some available information

Questions

Query ID, rank of the question n search result, title of question, snippet, body of the question

Status

Status of the question

Updated time

Last update time of the question

Answers

Number of the answers for the question, body of the best answer of the question

Views

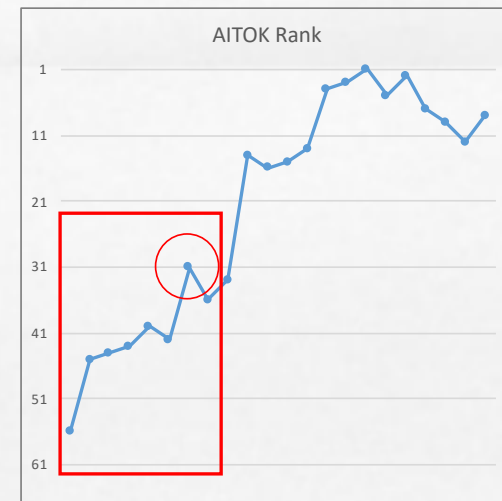Page view of the question

Clickthrough

most frequent rank of the question, Clickthrough rate

Where is my blue bird? (Q-score)

▨ In the offline test, we continued to submit a run a day from the end of August.
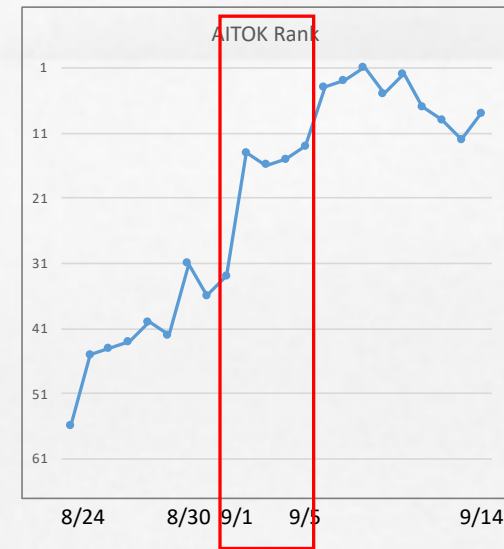
Date	Q-measure	Rank	Desc
8/24	0.38194	56	This result is only for uploading test from AITOK.
8/25	0.39724	45	1-gram TF-IDF with click through rate with cutoff
8/26	0.39852	44	1-gram TF-IDF+ with click through rate with cutoff
8/27	0.40479	43	2-gram TF-IDF+ with click through rate with cutoff
8/28	0.42008	40	2-gram TF-IDF+ with click through rate with cutoff without rank
8/29	0.41748	42	Dependent 2-gram TF-IDF with click through rate with cutoff without rank
8/30	0.4391	31	2-gram TF-IDF+ with click with cutoff and view without rank
8/31	0.42676	36	2-gram TF-IDF+ with click and view with cutoff without rank
9/1	0.43231	33	cutoff and view
9/2	0.49363	14	view count
9/3	0.49319	16	click through and view count
9/4	0.49347	15	view count sorted with click, updated, answers, order, rank and cutoff
9/5	0.49393	13	view count sorted with answers, cutoff, click, updated, order and rank
9/6	0.499	4	view count worted with answers x tf-idf weighted by query
9/7	0.5	3	view count + answers x 2-gram tf-idf weighted by query
9/8	0.50152	1	view count + answers x snippet 2-gram tf-idf weighted by query
9/9	0.49838	5	view count + answers x snippet 2-gram tf-idf weighted by query
9/10	0.50028	2	view count + answers x snippet 2-gram tf-idf double-weighted by norm query
9/11	0.49483	7	view count + answers x snippet word2vec double-weighted by norm query
9/12	0.49427	9	view count + answers x snippet word2vec double-weighted by norm query v2
9/13	0.49412	12	view count + answers x snippet L1 word2vec double-weighted by norm query
9/14	0.49437	8	view count + answers x snippet cos word2vec double-weighted by norm query

AITOK Rank

9

Where is my blue bird? (Q-score)

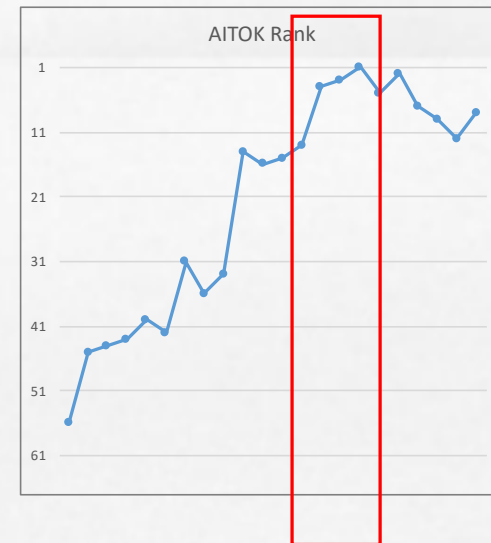▨ Which is important ? (view count, answers, click-through, update date, etc.)

Date	Q-measure	Rank	Desc
8/24	0.38194	56	This result is only for uploading test from AITOK.
8/25	0.39724	45	1-gram TF-IDF with click through rate with cutoff
8/26	0.39852	44	1-gram TF-IDF+ with click through rate with cutoff
8/27	0.40479	43	2-gram TF-IDF+ with click through rate with cutoff
8/28	0.42008	40	2-gram TF-IDF+ with click through rate with cutoff without rank
8/29	0.41748	42	Dependent 2-gram TF-IDF with click through rate with cutoff without rank
8/30	0.4391	31	2-gram TF-IDF+ with click with cutoff and view without rank
8/31	0.42676	36	2-gram TF-IDF+ with click and view with cutoff without rank
9/1	0.43231	33	cutoff and view
9/2	0.49363	14	view count
9/3	0.49319	16	click through and view count
9/4	0.49347	15	view count sorted with click, updated, answers, order, rank and cutoff
9/5	0.49393	13	view count sorted with answers, cutoff, click, updated, order and rank
9/6	0.499	4	view count worted with answers x tf-idf weighted by query
9/7	0.5	3	view count + answers x 2-gram tf-idf weighted by query
9/8	0.50152	1	view count + answers x snippet 2-gram tf-idf weighted by query
9/9	0.49838	5	view count + answers x snippet 2-gram tf-idf weighted by query
9/10	0.50028	2	view count + answers x snippet 2-gram tf-idf double-weighted by norm query
9/11	0.49483	7	view count + answers x snippet word2vec double-weighted by norm query
9/12	0.49427	9	view count + answers x snippet word2vec double-weighted by norm query v2
9/13	0.49412	12	view count + answers x snippet L1 word2vec double-weighted by norm query
9/14	0.49437	8	view count + answers x snippet cos word2vec double-weighted by norm query

AITOK Rank

| | 8/24 | 8/30 | 9/1 | 9/5 | 9/14 |

Where is my blue bird? (Q-score)

▨ View count and answers emphasizes the top score in offline test.

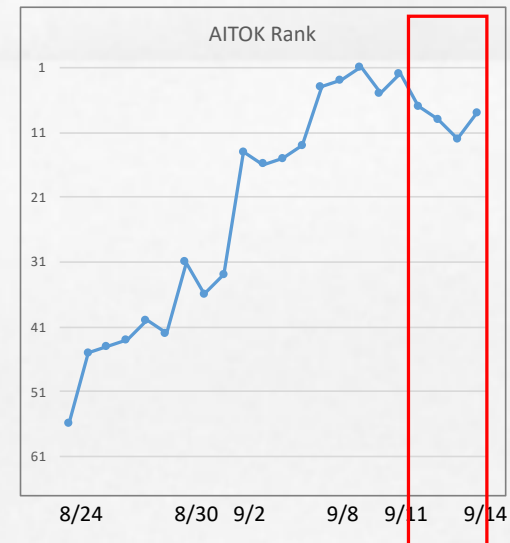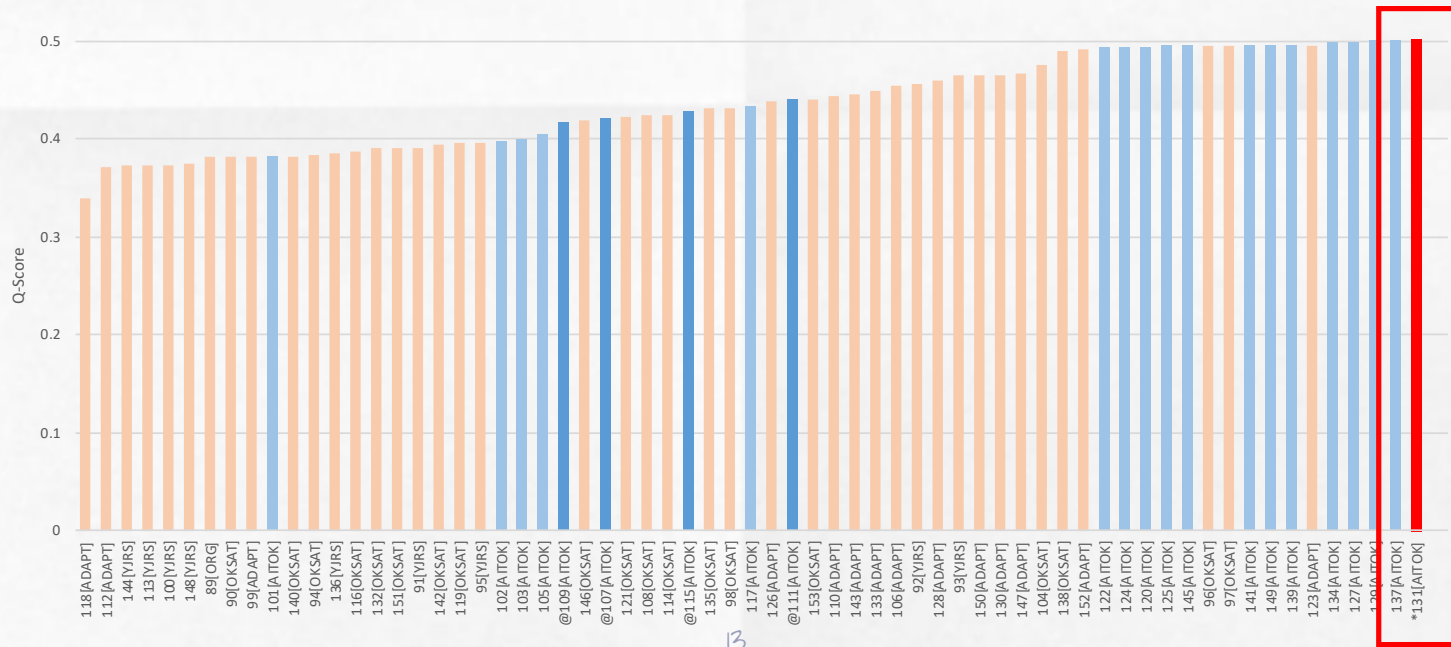| Date	Q-measure	Rank	Desc
8/24	0.38194	56	This result is only for uploading test from AITOK.
8/25	0.39724	45	1-gram TF-IDF with click through rate with cutoff
8/26	0.39852	44	1-gram TF-IDF+ with click through rate with cutoff
8/27	0.40479	43	2-gram TF-IDF+ with click through rate with cutoff
8/28	0.42008	40	2-gram TF-IDF+ with click through rate with cutoff without rank
8/29	0.41748	42	Dependent 2-gram TF-IDF with click through rate with cutoff without rank
8/30	0.4391	31	2-gram TF-IDF+ with click with cutoff and view without rank
8/31	0.42676	36	2-gram TF-IDF+ with click and view with cutoff without rank
9/1	0.43231	33	cutoff and view
9/2	0.49363	14	view count
9/3	0.49319	16	click through and view count
9/4	0.49347	15	view count sorted with click, updated, answers, order, rank and cutoff
9/5	0.49393	13	view count sorted with answers, cutoff, click, updated, order and rank
9/6	0.499	4	view count worted with answers x tf-idf weighted by query
9/7	0.5	3	view count + answers x 2-gram tf-idf weighted by query
9/8	0.50152	1	view count + answers x snippet 2-gram tf-idf weighted by query
9/9	0.49838	5	view count + answers x snippet 2-gram tf-idf weighted by query
9/10	0.50028	2	view count + answers x snippet 2-gram tf-idf double-weighted by norm query
9/11	0.49483	7	view count + answers x snippet word2vec double-weighted by norm query
9/12	0.49427	9	view count + answers x snippet word2vec double-weighted by norm query v2
9/13	0.49412	12	view count + answers x snippet L1 word2vec double-weighted by norm query
9/14	0.49437	8	view count + answers x snippet cos word2vec double-weighted by norm query

AITOK Rank

||

Where is my blue bird? (Q-score)

▨ I tried using word2vec.

Date	Q-measure	Rank	Desc
8/24	0.38194	56	This result is only for uploading test from AITOK.
8/25	0.39724	45	1-gram TF-IDF with click through rate with cutoff
8/26	0.39852	44	1-gram TF-IDF+ with click through rate with cutoff
8/27	0.40479	43	2-gram TF-IDF+ with click through rate with cutoff
8/28	0.42008	40	2-gram TF-IDF+ with click through rate with cutoff without rank
8/29	0.41748	42	Dependent 2-gram TF-IDF with click through rate with cutoff without rank
8/30	0.4391	31	2-gram TF-IDF+ with click with cutoff and view without rank
8/31	0.42676	36	2-gram TF-IDF+ with click and view with cutoff without rank
9/1	0.43231	33	cutoff and view
9/2	0.49363	14	view count
9/3	0.49319	16	click through and view count
9/4	0.49347	15	view count sorted with click, updated, answers, order, rank and cutoff
9/5	0.49393	13	view count sorted with answers, cutoff, click, updated, order and rank
9/6	0.499	4	view count worted with answers x tf-idf weighted by query
9/7	0.5	3	view count + answers x 2-gram tf-idf weighted by query
9/8	0.50152	1	view count + answers x snippet 2-gram tf-idf weighted by query
9/9	0.49838	5	view count + answers x snippet 2-gram tf-idf weighted by query
9/10	0.50028	2	view count + answers x snippet 2-gram tf-idf double-weighted by norm query
9/11	0.49483	7	view count + answers x snippet word2vec double-weighted by norm query
9/12	0.49427	9	view count + answers x snippet word2vec double-weighted by norm query v2
9/13	0.49412	12	view count + answers x snippet L1 word2vec double-weighted by norm query
9/14	0.49437	8	view count + answers x snippet cos word2vec double-weighted by norm query

AITOK Rank

12

Offline Test Result (Q-measure, nDGC, ERR)

▨ We won the first place in almost all scores.

My blue bird is here!

My blue bird, Catchy is hiding in the view count and TF-IDF :

- ▨ **Views**: the number of view, Page view of the question
- ▨ **Questions:** body and snippet with TF-IDF

Besides,

- ▨ Status: unconfirmed
- ▨ Updated time: unconfirmed
- ▨ Answers: a bit affected
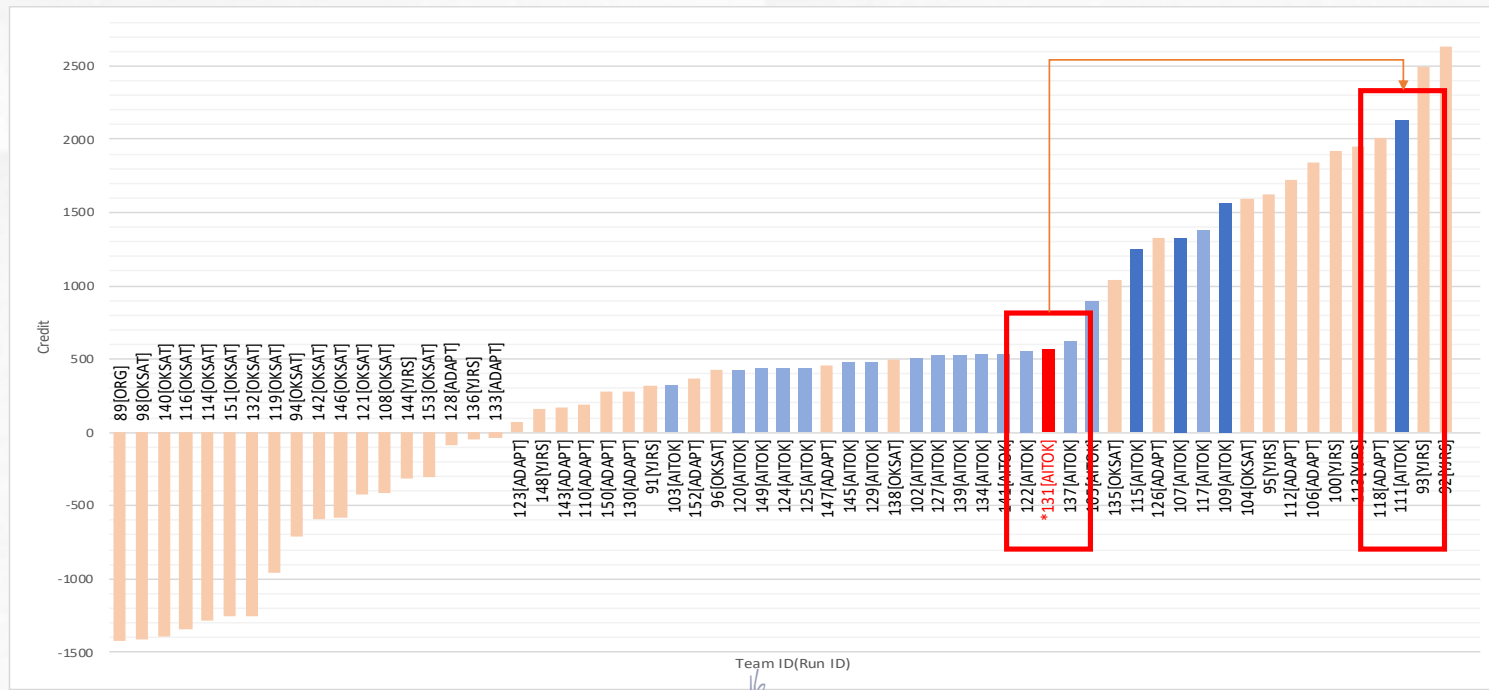- ▨ Clickthrough: a bit affected
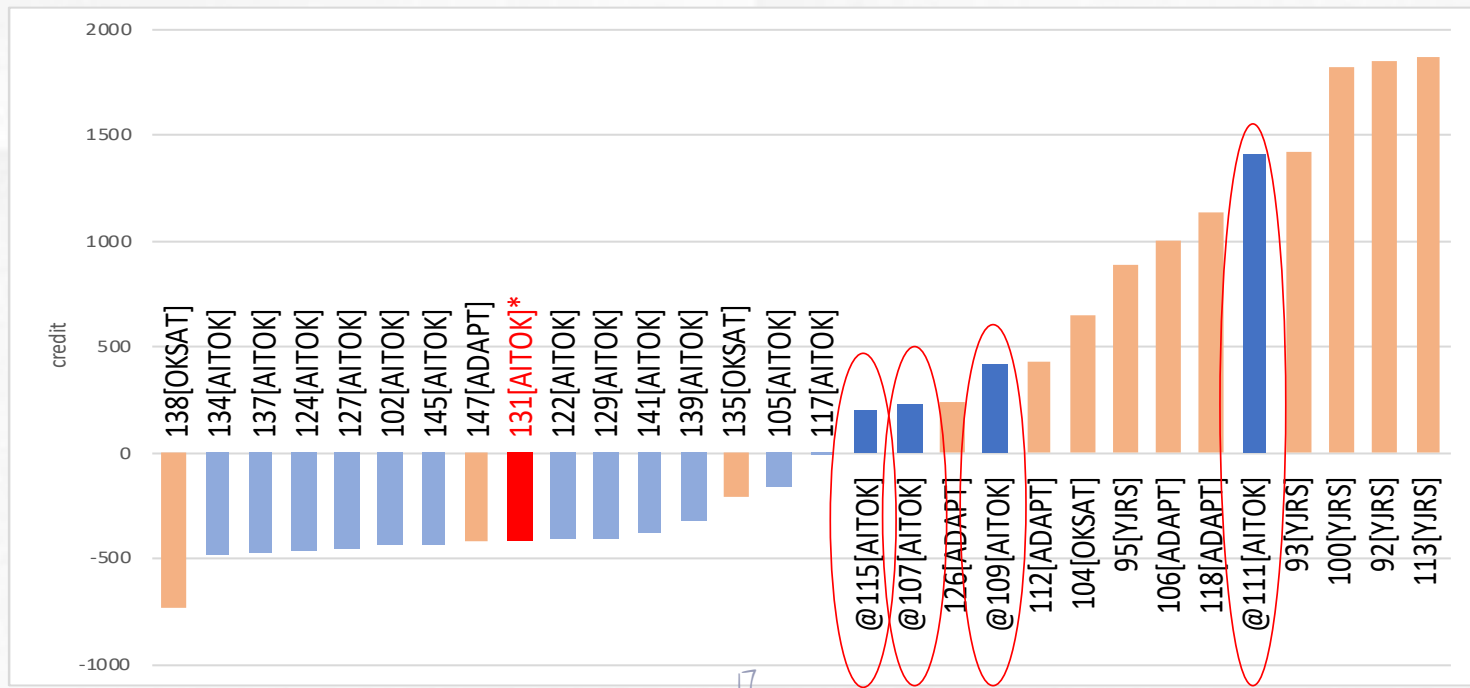
2.

Online Test

To the forest…

Online Test Result (top60 credit)

▨ ID131 is overtaken by ID111.

Online Test Result (top30 credit)

Where did my blue bird fly away?

What's happened?

ID131 was at the top of the offline test, but ID111 took over in the online test.

To the forest again… (top30 credit)

▨ What is ID111?

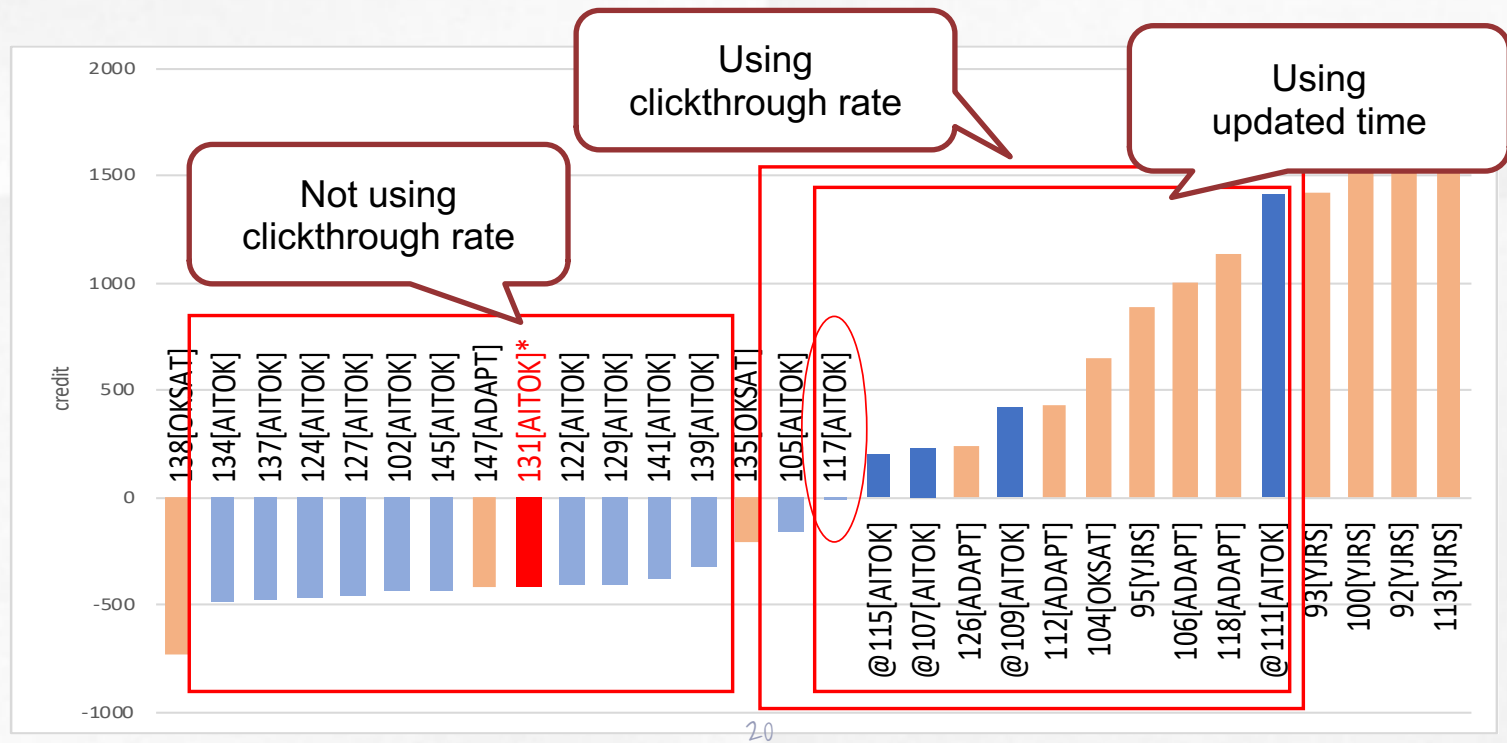| Date	Q-measure	Rank	Desc
8/24	0.38194	56	This result is only for uploading test from AITOK.
8/25	0.39724	45	1-gram TF-IDF with click through rate with cutoff
8/26	0.39852	44	1-gram TF-IDF+ with click through rate with cutoff
8/27	0.40479	43	2-gram TF-IDF+ with click through rate with cutoff
8/28	0.42008	40	2-gram TF-IDF+ with click through rate with cutoff without rank
8/29	0.41748	42	Dependent 2-gram TF-IDF with click through rate with cutoff without rank
8/30	0.4391	31	2-gram TF-IDF+ with click with cutoff and view without rank
8/31	0.42676	36	2-gram TF-IDF+ with click and view with cutoff without rank
9/1	0.43231	33	cutoff and view
9/2	0.49363	14	view count
9/3	0.49319	16	click through and view count
9/4	0.49347	15	view count sorted with click, updated, answers, order, rank and cutoff
9/5	0.49393	13	view count sorted with answers, cutoff, click, updated, order and rank
9/6	0.499	4	view co...
9/7	0.5	3	view co...
9/8	0.50152	1	view co...
9/9	0.49838	5	view co...
9/10	0.50028	2	view co...
9/11	0.49483	7	view count + answers x snippet word2vec double-weighted by norm query
9/12	0.49427	9	view count + answers x snippet word2vec double-weighted by norm query v2
9/13	0.49412	12	view count + answers x snippet L1 word2vec double-weighted by norm query
9/14	0.49437	8	view count + answers x snippet cos word2vec double-weighted by norm query

Using clickthrough rate

Using updated time

$$Score = cutoff \times (1 + \log \frac{views}{now - updated})$$

19

Comparison of Two Type of Tests

Offline

View count and TF-IDF based query search are effective in every metric at evaluation with relevance judgement data.

Online

Clickthrough rate and Updated time are effective in credit at evaluation with real users.

Summary

Offline test results were incredibly good:
- TF-IDF guesses intention of users.
- View count represents reputation.

Online test result were unexpected:
- Clickthrough are effective for live users.
- Updated time are also effective.

Want is Catchy?

Reputation
&
Freshness

Thanks!

Any questions?

You can find me at:
@taniokah
tanioka.hiroki@tokushima-u.ac.jp