

Results

Summary of bad results *[Updated]*

Union of all questions in which any of the scenarios performed poorly, ie. Recall@60 \leq 0.2 - *[Updated]*

- "q_id": Id of the question.
- Columns A, B, C, D: Value of Recall@60 for that scenario.
- Columns Q+/U+: Effect of the addition of more tags from questions/users compared against baseline
 - \uparrow/\downarrow : minor improvement/decay in the performance against the baseline (Recall(B/C/D) - Recall(A) \leq 0.4)
 - $\uparrow\uparrow/\downarrow\downarrow$: major improvement/decay in the performance against the baseline (Recall(B/C/D) - Recall(A) $>$ 0.4)
 - "-" : null effect in the performance against the baseline

q_id	A - Base	B - Q+	C - U+	D - Q+ U+	Q+	U+
9135	0.2	0.2	1.0	0.8	$\uparrow\uparrow$	$\uparrow\uparrow$
9197	0.8	0.6	0.6	0.4	\downarrow	\downarrow
9251	0.8	0.6	0.8	0.4	-	-
9273	0.2	0.2	0.6	0.4	\uparrow	\uparrow
9277	0.2	0.0	0.4	0.4	\uparrow	\uparrow
9427	0.4	0.2	1.0	0.6	$\uparrow\uparrow$	$\uparrow\uparrow$
9606	0.6	0.4	0.8	0.6	\uparrow	\uparrow
9615	0.2	0.0	0.4	0.4	\uparrow	\uparrow
9620	0.8	0.8	0.8	0.8	-	-
9653	0.4	0.2	0.2	0.2	\downarrow	\downarrow
9684	0.8	0.8	1.0	0.6	\uparrow	\uparrow
9750	0.4	0.4	0.8	0.4	\uparrow	\uparrow
9796	0.6	0.8	0.8	0.6	\uparrow	\uparrow

10006	0.2	0.2	0.2	0.0	-	-
10013	0.8	0.6	0.8	0.8	-	-
10020	0.4	0.2	0.4	0.4	-	-
10027	0.2	0.2	0.2	0.8	↑↑	↑↑
10072	0.4	0.4	0.4	0.8	↑	↑
10105	0.2	0.2	0.4	0.4	↑	↑
10120	0.2	0.2	0.2	0.2	-	-
10149	0.8	0.4	1.0	0.2	↑	↑
10238	0.4	0.0	0.4	0.2	-	-
10280	0.4	0.2	0.8	0.0	↑	↑
10284	0.2	0.2	0.2	0.2	-	-
10313	0.2	0.2	0.6	0.4	↑	↑
10323	0.4	0.4	0.6	0.4	↑	↑
10342	0.8	0.8	0.4	0.2	-	-

Summary of bad results: Questions with Recall@60 <= 0.2 in at least one for each scenario

Union of all questions in which any of the scenarios performed poorly, ie. Recall@60 <= 0.2

- "q_id": Id of the question.
- Columns A, B, C, D: Value of Recall@60 for that scenario.
- Columns Q+/U+: Effect of the addition of more tags from questions/users compared against baseline
 - ↑/↓ : minor improvement/decay in the performance against the baseline (Recall(B/C/D) - Recall(A) <= 0.4)
 - ↑↑/↓↓ : major improvement/decay in the performance against the baseline (Recall(B/C/D) - Recall(A) > 0.4)
 - "-" : null effect in the performance against the baseline

q_id	A - Base	B - Q+	C - U+	D - Q+ U+	Q+	U+

9135	0.2	0.8	0.2	0.8	↑↑	-
9197	0.8	0.2	0.8	0.2	↓↓	-
9251	0.8	0.2	0.8	0.2	↓↓	-
9273	0.2	0.4	0.2	0.4	↑	-
9277	0.2	0.2	0.2	0.2	-	-
9427	0.4	0.2	0.4	0.2	↓	-
9606	0.6	0.2	0.6	0.2	↓	-
9615	0.2	0.2	0.2	0.2	-	-
9620	0.8	0.2	0.8	0.2	↓↓	-
9653	0.4	0.0	0.4	0.0	↓↓	-
9684	0.8	0.2	0.8	0.2	↓↓	-
9750	0.4	0.0	0.4	0.0	↓	-
9796	0.6	0.0	0.6	0.0	↓↓	-
10006	0.2	0.0	0.2	0.0	↓	-
10013	0.8	0.0	0.8	0.0	↓↓	-
10020	0.4	0.0	0.4	0.0	↓	-
10027	0.2	0.6	0.2	1.0	↑	↑
10072	0.4	0.2	0.4	0.2	↓	-
10105	0.2	0.0	0.2	0.0	↓	-
10120	0.2	0.8	0.2	0.8	↑	-
10149	0.8	0.0	0.8	0.0	↓↓	-
10238	0.4	0.2	0.4	0.2	↓	-
10280	0.4	0.0	0.6	0.0	↓↓	↑
10284	0.2	0.0	0.2	0.0	↓	-
10313	0.2	0.0	0.2	0.0	↓	-
10323	0.4	0.0	0.4	0.0	↓	-
10342	0.8	0.0	0.8	0.0	↓↓	-

Questions with worse recall for each scenario

We show the questions with a recall `<=0.2` counting the top 60 best ranked users. The script for this is in the file: `reports/recall_diff.r`

Scenario A - Baseline

q_id	top.15	top.30	top.45	top.60
9273	0.0	0.0	0.0	0.2
9277	0.0	0.0	0.0	0.2
10006	0.0	0.2	0.2	0.2
9135	0.2	0.2	0.2	0.2
9615	0.2	0.2	0.2	0.2
10027	0.2	0.2	0.2	0.2
10105	0.2	0.2	0.2	0.2
10120	0.2	0.2	0.2	0.2
10284	0.2	0.2	0.2	0.2
10313	0.2	0.2	0.2	0.2

Scenario B - Questions' tags description extended

q_id	top.15	top.30	top.45	top.60
9653	0.0	0.0	0.0	0.0
9750	0.0	0.0	0.0	0.0
9796	0.0	0.0	0.0	0.0
10006	0.0	0.0	0.0	0.0
10013	0.0	0.0	0.0	0.0
10020	0.0	0.0	0.0	0.0
10105	0.0	0.0	0.0	0.0
10149	0.0	0.0	0.0	0.0
10280	0.0	0.0	0.0	0.0
10284	0.0	0.0	0.0	0.0
10313	0.0	0.0	0.0	0.0
10323	0.0	0.0	0.0	0.0
10342	0.0	0.0	0.0	0.0
9277	0.0	0.0	0.0	0.2
10238	0.0	0.0	0.0	0.2
9620	0.0	0.0	0.2	0.2
9197	0.0	0.2	0.2	0.2
9251	0.0	0.2	0.2	0.2
9615	0.0	0.2	0.2	0.2
9684	0.0	0.2	0.2	0.2
9427	0.2	0.2	0.2	0.2
9606	0.2	0.2	0.2	0.2
10072	0.2	0.2	0.2	0.2

Scenario C - Users' tags description extended

q_id	top.15	top.30	top.45	top.60
9273	0.0	0.0	0.0	0.2
9277	0.0	0.0	0.0	0.2
10006	0.0	0.2	0.2	0.2
9135	0.2	0.2	0.2	0.2
9615	0.2	0.2	0.2	0.2
10027	0.2	0.2	0.2	0.2
10105	0.2	0.2	0.2	0.2
10120	0.2	0.2	0.2	0.2
10284	0.2	0.2	0.2	0.2
10313	0.2	0.2	0.2	0.2

Scenario D - Questions' AND Users' tags description extended

q_id	top.15	top.30	top.45	top.60
9653	0.0	0.0	0.0	0.0
9750	0.0	0.0	0.0	0.0
9796	0.0	0.0	0.0	0.0
10006	0.0	0.0	0.0	0.0
10013	0.0	0.0	0.0	0.0
10020	0.0	0.0	0.0	0.0
10105	0.0	0.0	0.0	0.0
10149	0.0	0.0	0.0	0.0
10280	0.0	0.0	0.0	0.0
10284	0.0	0.0	0.0	0.0
10313	0.0	0.0	0.0	0.0
10323	0.0	0.0	0.0	0.0
10342	0.0	0.0	0.0	0.0
9277	0.0	0.0	0.0	0.2
10238	0.0	0.0	0.0	0.2
9620	0.0	0.0	0.2	0.2
9197	0.0	0.2	0.2	0.2
9251	0.0	0.2	0.2	0.2
9615	0.0	0.2	0.2	0.2
9684	0.0	0.2	0.2	0.2
9427	0.2	0.2	0.2	0.2
9606	0.2	0.2	0.2	0.2
10072	0.2	0.2	0.2	0.2

Second experiments (TMBA): using extended tag description of questions and users

Objective

To determine if describing questions and/or users with the tags in the body and title of questions can improve the behaviour of TMBA.

Scenarios

A. **Baseline**: no use of extended tag description B. **Questions extended** tag description but Users non-extended C. Questions non-extended and **Users extended** tag description D. **Questions extended** tag description - **Users extended** tag description

Tasks

- ☒ Extract new tags from Questions: generate `ros_question_tag_extended.json`
- ☒ Extract new tags from participation of Users in Questions: generate `ros_user_tag_extended.json` --> (`ra_user_tag_extended` in the DB)
- ☒ Generate R_ut for Scenario B: generate `r_ut_scenario_b.json`
- ☒ Generate R_ut for Scenario C: generate `r_ut_scenario_c.json`
- ☒ Generate R_ut for Scenario D: generate `r_ut_scenario_d.json`
- ☒ Get scores for Scenario B
- ☒ Get scores for Scenario C
- ☒ Get scores for Scenario D
- ☒ Overall evaluation for Scenario B (boxplot of recall)
- ☒ Overall evaluation for Scenario C (boxplot of recall)
- ☒ Overall evaluation for Scenario D (boxplot of recall)
- ☐ Manual inspection

First experiments: WCFA/C2AA/TMBA/Karma

Initial exploratory experiments for testing WCFA, TMBA, C2AA against the RPK (baseline).

Method	Nb of Questions	Nb of Participants	Folder	Observations
WCFA	100	1 asker + 4 answerers	wcfa_100q_5p	See the results .
C2AA	100	1 asker + 4 answerers	wcfa_100q_5p_4hidden	This method is the same as WCFA but for R_{uu} there is no knowledge about participation in the target question. See the results .
TMBA	100	1 asker + 4 answerers	tmba_100q_5p	See the results .
TMBA	100	1 asker + 0 answerers	tmba_100q_1p	TMBA for the COLD START problem. See the results .
RPK	100	1 asker + 4 answerers	rpk_100p_5p	PseudoKarma. See the results .