

Приложение I.

Результаты дорожки вопросно-ответного поиска.

В этом приложении представлена сводная информация о результатах оценки для дорожки вопросно-ответного поиска.

Асессоры производили оценку по шкале:

1. Точный ответ
2. Частичный ответ
3. Ответа нет, но из фрагмента похоже, что он рядом
4. Нет ответа

При вычислении приведенных ниже оценок считалось, что ответ «правильный», если оценка асессора 1 или 2.

На момент публикации трудов из-за задержки с получением второй оценки оценки асессоров не были объединены. Отметим, что для 226 вопросов (из 615) по мнению первого асессора не нашлось ни одного правильного ответа (для 146 из них не было даже ни одного ответа с оценкой 3), по мнению второго асессора таких вопросов было 231 (и 271).

В приведенных ниже таблицах используются следующие метрики:

- **TotalAnswers:** общее число ответов возвращенных системой для всем 615 Запросов.
- **TotalRelevant:** общее число ответов получивших оценку 1 или 2.
- **TotalWeakRelevant:** общее число ответов получивших оценку 3.
- **Precision (micro)** – микроусреднение точности на уровне 10 по всем заданиям.
- **Precision (macro)** - макроусреднение точности на уровне 10 по всем заданиям.
- **TrecReciprocalRank:** усредненная по всем вопросам «ценность» ответов с использованием линейки TREC (1.0, 0.5, 0.33, 0.2, 0.1)
- **ROMIPReciprocalRank:** усредненная по всем вопросам «ценность» ответов с использованием линейки РОМИП (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1)

Более подробно метрики описаны в приложении А.

Значения метрик вычисленные по оценкам первого ассессора (вычислены по всем 615 заданиям):

Метрика	Прогон 1	Прогон 2
TotalAnswers	5509	6145
TotalRelevant	880	456
TotalWeakRelevant	158	280
Precision (micro)	0.1597	0.074
Precision (macro)	0.1744	0.074
TrecReciprocalRank	0.410	0.270
ROMIPReciprocalRank	0.483	0.343

Значения метрик вычисленные по оценкам второго ассессора (расчитаны по 426 заданиям, не учтенные 189 вопросов не содержали ни одного правильного ответа):

Метрика	Прогон 1	Прогон 2
TotalAnswers	3939	4245
TotalRelevant	705	394
TotalWeakRelevant	135	113
Precision (micro)	0.1789	0.092
Precision (macro)	0.2044	0.092
TrecReciprocalRank	0.492	0.319
ROMIPReciprocalRank	0.600	0.417