

Результаты соревнований Диалог-2016 по выделению именованных сущностей (FactRuEval, Dialogue Evaluation 2016)

В 2016 году в рамках Dialogue Evaluation проводилось соревнование по извлечению информации из новостных текстов на русском языке (см. <http://www.dialog-21.ru/evaluation/2016/ner/>). Были предложены 3 дорожки, на каждой по 2 «поддорожки»:

- **T1**: определение в тексте границ таких именованных сущностей, как персона (PER), организация (ORG), локация (LOC), а также некая «орголокация» (LOCORG), к которой не очень внятно относятся организации (когда они вроде как локации) и локации (когда они вроде как не локации)¹;
- **T1_I**: модификация T1, когда считается, что LOCORG=LOC;
- **T2**: определение и нормализация атрибутов вышеупомянутых именованных сущностей, сущности не должны дублироваться (то есть необходимо объединять по сути одинаковые сущности);
- **T2_m**: то же, что T2, но лишние атрибуты не штрафуются;
- **T3, T3_m**: выделение фактов (мы здесь не участвовали).

Сначала участникам предоставили 122 размеченных текста для настройки, затем на 2 недели дали боевые тексты в количестве около 37000. Участники могли присылать несколько прогонов.

Участники выступали под псевдонимами, Pullenti был под псевдонимом **Pink**.

Результаты оценивались по так называемой ф-мере $= 2 * P * R / (P + R)$, где P – точность, R – полнота. **Pullenti** показал следующий результат:

T1 – **первое место** (13 участников);

T1_I – **второе место** (13 участников);

T2, T2_m – **первое место** (5 участников);

T3, T3_m – не участвовал (2 участника).

Участник	T1	T1_I	T2	T2_m
Pink (Pullenti)	0.8088	0.8575	0.7979	0.8126
Aquamarine (Promt)	0.8069	0.8475	0.7916	0.7938
Violet (Compreno)	0.7682	0.8672	0.7819	0.7857
Crimson	0.6908	0.8493	0.7772	0.7799
Beige	0.7022	0.8558	-	-
Gray (RCO)	0.4353	0.4356	-	-
...				

¹ Например, «презентация состоялась в магазине Библио-Глобус» - это именно LOCORG с точки зрения организаторов. «состоялось в России» - это LOC, а «Россия заявила...», «президент России» - это LOCORG.

То есть Pullenti обошёл только участник Violet (Compreno, Abbyu) на дорожке T1_I.

Отметим, что RCO (Gray) принимал ограниченное участие, выделяя только персоны. Хотя суммарная ф-мера низка, по персонам она = 0.912 . У нас по персонам 0.913.

Вот диаграммы по дорожкам. Маленькие фигурки отображают результаты на демо-выборке из 122 текстов, большие – на тестовой выборке. Круг – ф-мера, квадрат – точность, треугольник – полнота. Цифрой обозначается номер прогона (можно было несколько).



