

В 2017 году мы приняли участие в соревновании MorphRuEval, проводимое в рамках ежегодной конференции Диалог (см. <http://www.dialog-21.ru/evaluation/2017/morphorueval/>). Участникам предлагались 2 дорожки: без лемматизации (только морфологические признаки) и с лемматизацией.

Таблица с результатами пока находится здесь:

https://docs.google.com/spreadsheets/d/1npLGlvfxtjRiLRuQjd1rkbnr-nIKBWC_yKd_OnRYTnE/edit#gid=0

На дорожке с лемматизацией мы заняли 3-е место из 7, без лемматизации – 6 из 12. Вот результат с лемматизацией:

1 место: IQMEN – 92.22%

2 место: Аспект (НИИ Квант) – 91.81%

3 место: Pullenti POS Tagger – 89.32%

Результат довольно скромный, однако до этого мы морфологию фактически не развивали, вполне обходясь своим POS-Tagger-ом при решении практических задач. POS-Tagger отличается от полноценной морфологии тем, что предлагает для словоформы всевозможные морфологические варианты независимо от окружения этой словоформы, а морфология старается выбрать один вариант исходя из контекста словоформы.

Специально для конкурса был разработан стенд, в котором задавались правила для уменьшения морфологических вариантов словоформ. Правило оставлялось, если оно увеличивало суммарный коэффициент совпадений с эталонной разметкой. В день удавалось улучшить качество примерно на 1%. Не хватило каких-то нескольких дней...