

Домашнее задание №12

Асланов А.Б., ИУ9-11М

- Протестировать извлечения именованных сущностей на русском языке (на сайте **demo.ipavlov.ai**).
- Написать отчет: что хорошо, что плохо

На сайте сущности маркируются разными цветами:

Сущности: **Человек** **Организация** **Локация**

Больше всего текстов, где есть одновременно все три сущности - в новостном корпусе, поэтому возьмём примеры оттуда:

Пример 1.

Первый заместитель министра обороны **США** **Патрик Шанахан** с 1 января 2019 года станет исполняющим обязанности главы **Пентагона** после ухода с этой должности **Джеймса Мэттиса** . Об этом сообщил в воскресенье в своем **Twitter** президент страны **Дональд Трамп** . `` Рад объявить , что талантливый первый заместитель министра обороны **Патрик Шанахан** станет исполняющим обязанности главы **Пентагона** с 1 января 2019 года . У **Патрика** за плечами большой послужной список , он многого добился за время работы заместителем [министра] , а до этого - в компании **Boeing "** , - написал президент **США** . **Патрик Шанахан** родился и вырос с штате **Вашингтон** , окончил **Университет штата Вашингтон** по специальности машиностроение , в настоящее время он входит в попечительский совет этого вуза . **Шанахан** разведен , у него трое детей .

Все сущности определены верно.

Пример 2.

Пресс-секретарь президента **Дмитрий Песков** заявил в эфире « **Россия 1** » , что многие вопросы к **Путину** на пресс-конференции были жалобами , причем подданными гражданами , а не профессиональными журналистами . Все эти жалобы отнимали время у президента и других журналистов . Кроме того , после них **Путин** вынужден давать специальные поручения чиновникам . Среди подобных обращений есть те , которые являются клеветой , отметил **Песков** . Пресс-секретарь **Путина** пообещал подумать , как он будет бороться с подобной практикой .

Не была включена единичка в названии телеканала "Россия 1".

Пример 3.

Глава администрации **Клинцов** **Александр Морозов** сообщил , что замглавы горадминистрации по социальным вопросам , которая занималась отправкой группы детей на бесплатный отдых в **Турцию** , уволилась , передает **ТАСС** . Также , по его словам , начальник отдела образования получила выговор после проверки прокуратуры . « Проводилась прокурорская проверка , было представление в адрес главы администрации , по результатам рассмотрения которого был вынесен выговор начальнику отдела образования , а заместитель , курирующий социальные вопросы и занимавшийся формированием группы , ушла в отставку » , — сказал **Морозов** . Он также заявил , что замечания прокуратуры были связаны с недостаточным информированием населения о поездке в **Турцию** . Ранее стало известно , что детям брянских чиновников дали бесплатные путевки в **Турцию** . При этом финансировал поездку благотворительный фонд , занимающийся помощью детям с тяжелыми заболеваниями и оказавшимся в тяжелой жизненной ситуации .

Все сущности определены верно.

Пример 4.

Приглашённый профессор **Военного колледжа армии** **США** **Роберт Фарли** назвал регионы мира , в которых может начаться мировая война , в статье в **The National Interest** . Американский военный эксперт считает , что наибольшим конфликтным потенциалом обладают следующие регионы : **Южно-Китайское море** . Этот регион стал точкой столкновения интересов **Китая** и **США** . Пока стороны конфликта ограничиваются торговыми санкциями , однако **США** пошли на обострение , арестовав руководителя китайской технологической фирмы **Huawei** . Если отношения между **США** и **Китаем** продолжат ухудшаться , то именно **Южно-Китайское море** станет ареной конфликта . **Украина** . **Роберт Фарли** отметил , что инцидент в **Керченском проливе** заставил мир вспомнить об **Украине** . Он предположил , что предстоящие президентские выборы на **Украине** способны нарушить баланс сил , сформировавшийся в регионе , что может привести к хаосу . Персидский залив . В этом регионе конфликт может начаться сразу в нескольких странах . Одной из них является **Иран** , находящийся под тяжёлыми экономическими санкциями со стороны **США** . Также конфликт может разгореться и между **Саудовской Аравией** и **Йеменом** . Третьей конфликтной точкой остаётся **Сирия** , в которой возможно возобновление активных военных действий между курдами , турками , сирийцами и иракцами .

Южно-китайское море отмечено не как локация, а как организация. В остальном всё в порядке.

Пример 5.

Представители стран , поддержавших в **ООН** украинскую резолюцию о `` нарушениях прав человека в **Крыму** " , для начала должны побывать в республике , заявил первый вице-спикер крымского парламента **Ефим Фикс** . По словам парламентария , ему `` стыдно за те государства , которые голосуют за резолюцию , не зная сути вопроса " , передает **ТАСС** . Власти **Крыма** готовы принять комиссию **ООН** и допустить ее во все населенные пункты , не составляя для ее членов никаких программ , отметил он . И если представители **ООН** не обнаружат нарушений прав человека , они должны выйти на площадь имени **Ленина** в центре **Симферополя** и `` при многочисленном стечении народа извиниться за те безобразия , которые они творят вместе с **Украиной "** . Генеральная Ассамблея **ООН** большинством голосов приняла украинский вариант резолюции по **Крыму** , осуждающей `` нарушение прав человека " и `` незаконное установление российских законов , юрисдикции и управления " на полуострове . Документ поддержали 67 делегаций стран-членов **ООН** , в том числе **США** , **ЕС** , **Турция** и **Япония** . Против выступили 27 делегаций , включая представителей **Китая** , **Индии** , **Сирии** и **ЮАР** . Еще 70 стран при голосовании воздержались . Ранее замминистра иностранных дел **Украины** **Сергей Кислица** , выступая в **ООН** , обвинил российские власти в `` попытках , сексуальном насилии и других нарушениях прав жителей **Крыма** " . Заместитель постпреда **России** при **ООН** **Геннадий Кузьмин** в связи с этим подчеркнул , что **Украина** уже четыре года подряд упорно представляет мировому сообществу ситуацию в **Крыму** как вооруженный конфликт , тогда как полуостров перешел под юрисдикцию **РФ** в результате референдума . Глава республики **Сергей Аксенов** также пригласил представителей **ООН** , **ОБСЕ** , **Европарламента** в **Крым** , чтобы оценить реальную ситуацию в регионе . `` Подобные документы откровенно игнорируют мнение самих крымчан " , — считает **Аксенов** .

На шестой строчке Украину (в контексте имелись в виду жители Украины) алгоритм пометил как организацию. Также неверно была распознана *площадь имени Ленина*: это место-словочетание было помечено не целиком.

Пример 6

Попробуем взять художественную литературу, произведение И.А. Гончарова "Обломов":

— Нет , я еще отроду барского не пропивал ! — захрипел **Захар** . — Вот вы ... — Перестань , **Захар** ! — строго перебил **Обломов** . — Вы , что ли , увезли одну половую щетку да две чашки у нас ? — спросил опять **Захар** . — Какие щетки ? — загремел **Тарантьев** . — Ах ты , старая шельма ! Давай-ка лучше закуски ! — Слышите , **Илья Ильич** , как лается ? — сказал **Захар** . — Нет закуски , даже хлеба нет дома , и **Анисья** со двора ушла , — договорил он и ушел . — Где ж ты обедаешь ? — спросил **Тарантьев** . — Диво , право : **Обломов** гуляет в роще , не обедает дома ... Когда ж ты на квартиру-то ? Ведь осень на дворе . Приезжай посмотреть . — Хорошо , хорошо , на днях ... — Да деньги не забудь привезти ! — Да , да , да ... — нетерпеливо говорил **Обломов** . — Ну , не нужно ли чего на квартире ? Там , брат , для тебя выкрасили полы и потолки , окна , двери — все : больше ста рублей стоит .

Все сущности распознаны верно.

Выводы

- Плохо распознается метонимия;
- Не всегда хорошо распознаются локации, названия которых состоят более чем из одного слова;
- В остальном именованные сущности в целом извлекаются корректно, в том числе довольно сложные, будь то фамилии и организации.